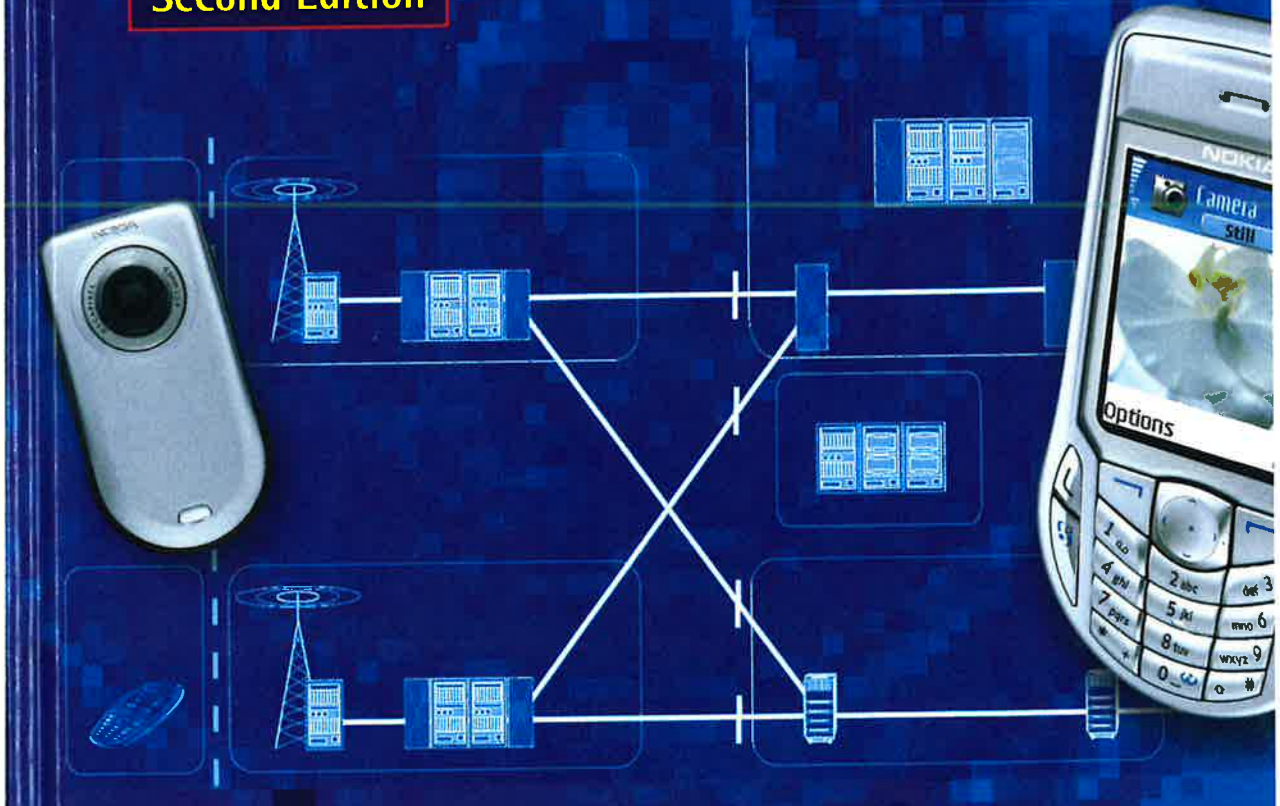
WILEY

UMTS NETWORKS

Architecture, Mobility and Services

Second Edition



Heikki Kaaranen · Ari Ahtiainen · Lauri Laitinen · Siamäk Naghian · Valtteri Niemi

UMTS Networks

Architecture, Mobility and Services

Second Edition

Heikki Kaaranen
Oy Aqua Records Ltd, Finland

Ari Ahtiainen
Nokia Research Center, Finland

Lauri Laitinen
Nokia Research Center, Finland

Siamäk Naghian
Nokia Networks, Finland

Valtteri Niemi
Nokia Research Center, Finland

JOHN WILEY & SONS, LTD

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

Reprinted June and December 2005, March 2006

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-10 0-470-01103-3 (H/B)
ISBN-13 978-0-470-01103-4 (H/B)

Project management by Originator, Gt Yarmouth, Norfolk (typeset in 10/12pt Times).
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire.
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

6

UMTS Core Network

Heikki Kaaranen and Miikka Poikselkä

The Universal Mobile Telecommunication System (UMTS) Core Network (CN) can be seen as the basic platform for all communication services provided to UMTS subscribers. The basic communication services include switching of circuit-switched calls and routing of packet data. The 3G Partnership Project (3GPP) R5 also introduces a new subsystem called the "IP Multimedia Subsystem" (IMS). The IMS opens up the Internet Protocol (IP)-based service world for mobile use by seamlessly integrating the mobile world and the Internet world and providing sophisticated service mechanisms to be used in the context of mobile communications.

The CN maps end to end Quality of Service (QoS) requirements to the UMTS bearer service. When inter-connecting with other networks, QoS requirements also need to be mapped onto the available external bearer service. The gateway role of the UMTS CN in creating an end-to-end service path is illustrated in Figure 6.1. The external bearer does not fall within the scope of UMTS system specifications and this may create some local problems if the QoS requirements to be satisfied between the UMTS and external network do not match.

Between the Mobile Termination (MT) and the CN the QoS is provided by the radio access bearer. The radio access bearer hides QoS handling over the radio path from the CN. Within the CN, QoS requirements are mapped to its own bearer service, which in turn is carried by backbone bearers on top of the underlying physical bearer service. A challenge to CN implementation is that the operator is pretty much free to choose how to implement physical backbone bearers. These bearers rely on the physical transmission technologies used between CN nodes. Typical transmission technologies, like PDH and SDH, with Pulse Code Modulation (PCM) channelling or with Asynchronous Transfer Mode (ATM) cell-switching are used. In 3GPP R5 the emphasis is on replacing these technologies by the Internet Protocol (IP) wherever and whenever possible, since making this transport network uniform simplifies the functionality of higher protocol layers.

The UMTS represents a kind of philosophy for use in production of a universal core that is able to handle a wide set of different radio accesses. Looking back at the network evolution discussed in Chapter 2, we see there are three types of recognised radio accesses as far as 3GPP R5 is concerned: WCDMA/HSDPA, GSM/EDGE and,

UMTS Networks Second Edition H. Kaaranen, A. Ahtiainen, L. Laitinen, S. Naghian and V. Niemi
© 2005 John Wiley & Sons, Ltd ISBN: 0-470-01103-3

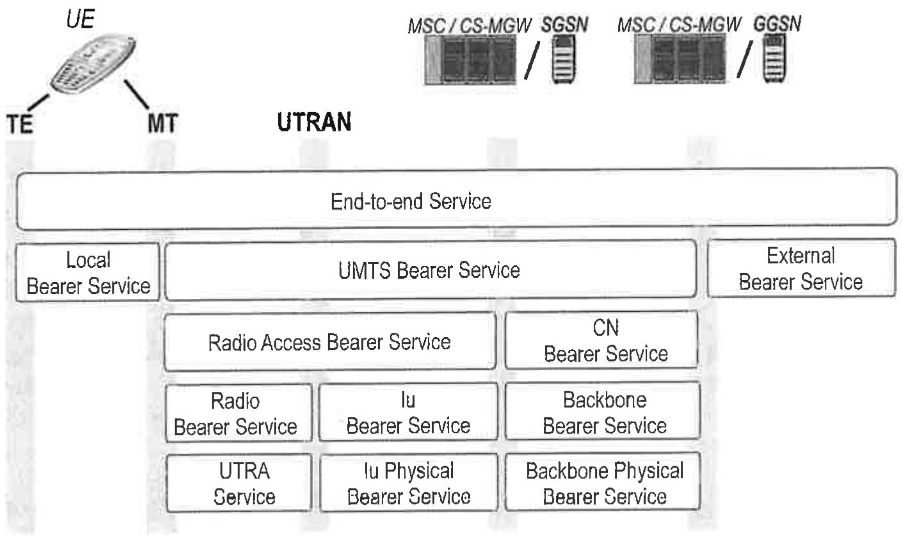


Figure 6.1 Bearer and Quality of Service (QoS) architecture in the Core Network (CN)

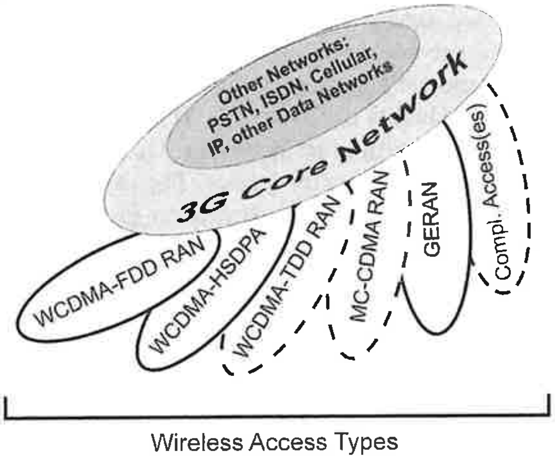


Figure 6.2 Universal core for wireless access

possibly, complementary access. Of these, WCDMA/HSPDA and GSM/EDGE are implemented, while complementary access is under study. The core part of the UMTS network does not evolve in as straightforward a way as the radio network due to both the CN's traditional infrastructure basis and its advanced technologies, which may have a number of different impacts on the evolution of the core part of the UMTS. Figure 6.2 shows the conceptual nature of the UMTS CN: the radio accesses drawn as continuous lines are the ones used at the outset and the others are regarded as access candidates as time goes by.

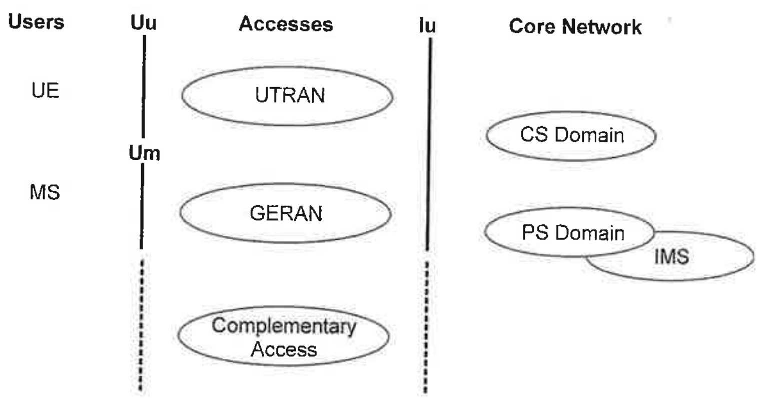


Figure 6.3 Core Network (CN) structure on a domain/subsystem level

6.1 UMTS Core Network Architecture

3GPP R99 introduced new mechanisms and capacity increases for the access network side. Starting with 3GPP R4 and its actual realisation in 3GPP R5 the CN has undergone major changes. In this chapter we will introduce, albeit briefly, the main characteristics of 3GPP R5.

As shown in Figure 6.3 the UMTS CN consists of equipment entities called "domains" and "subsystems" whose purpose is to describe the traffic characteristics the equipment takes care of. Based on this division, the UMTS CN contains the following entities:

- Circuit Switched (CS) domain.
- Packet Switched (PS) domain.
- IP Multimedia Subsystem (IMS).
- BroadCast (BC) domain.

What is the difference between a domain and a subsystem as far as the CN is concerned? The CN domain is an entity directly interfacing one or more access networks. This interface is called "Iu". Due to the nature of traffic and to identify the domain, the Iu is very often subscripted: for example, Iu_{CS} is the interface between an access network and the CS domain and delivers CS traffic; Iu_{PS} is the interface for PS traffic purposes; and Iu_{BC} is the interface that carries a broadcast/multicast type of traffic. CN subsystems do not have a direct Iu-type interface with access networks. Instead, they utilise other, separately defined interfaces to connect themselves to one or more CN domains.

Figure 6.4 is not exhaustive but aims to illustrate the most important interfaces within the UMTS CN. For a complete presentation, see 3GPP TS23.002, version 5.12.0.

In this figure, the bold lines indicate user traffic (user plane) and thinner lines indicate signalling connections (control plane). As far as the CN is concerned, there are some items that need to be pointed out:

- The connections drawn in the figure represent logical, direct connections. In reality, however, the connections have other ways of connecting due to transport network solutions.

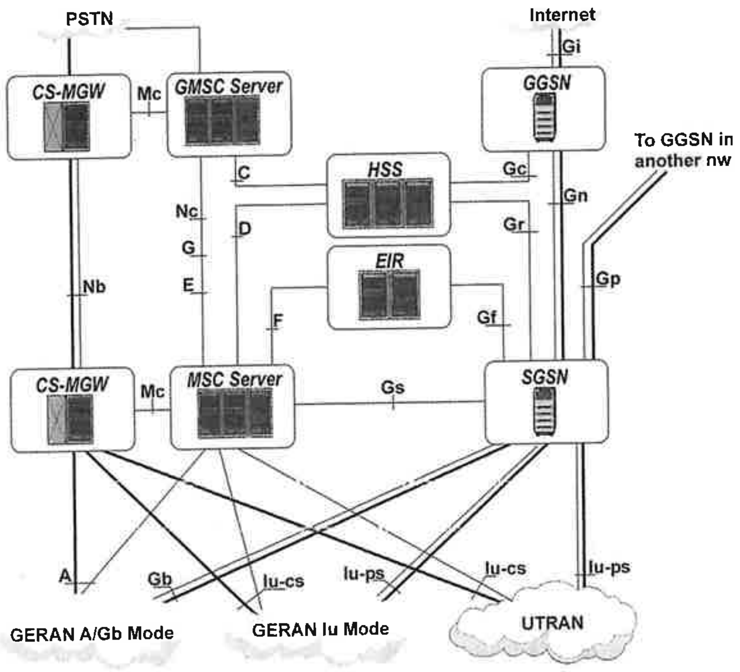


Figure 6.4 Core network (CN) configuration supporting Circuit Switched (CS) and Packet Switched (PS) traffic

- The CS Media Gateway (CS-MGW) and the Gateway Mobile Services Switching Centre (GMSC) server can be combined into one physical entity. In this case the entity is simply called the "GMSC".
- If the CS domain structure follows 3GPP R99, the CS-MGW and MSC Server could be combined into one physical entity. In this case the entity is called the MSC/VLR (Visitor Location Register).
- If the Serving GPRS Support Node (SGSN) and MSC/VLR are combined into one physical entity it is called the UMSC (UMTS MSC).

Sections 6.1.1–6.1.3 handle CN-domain-related issues. The IMS is handled in Sections 6.4–6.6.

CN management tasks and control duties with related issues, like identities and addressing, are handled in Section 6.2.

6.1.1 Core Network Entities that Are Common to All Domains and Subsystems

In addition to domains the CN contains some functionalities that are common to all CN domains and subsystems. These common functionalities are mainly collected in an entity called the "Home Subscriber Server" (HSS).

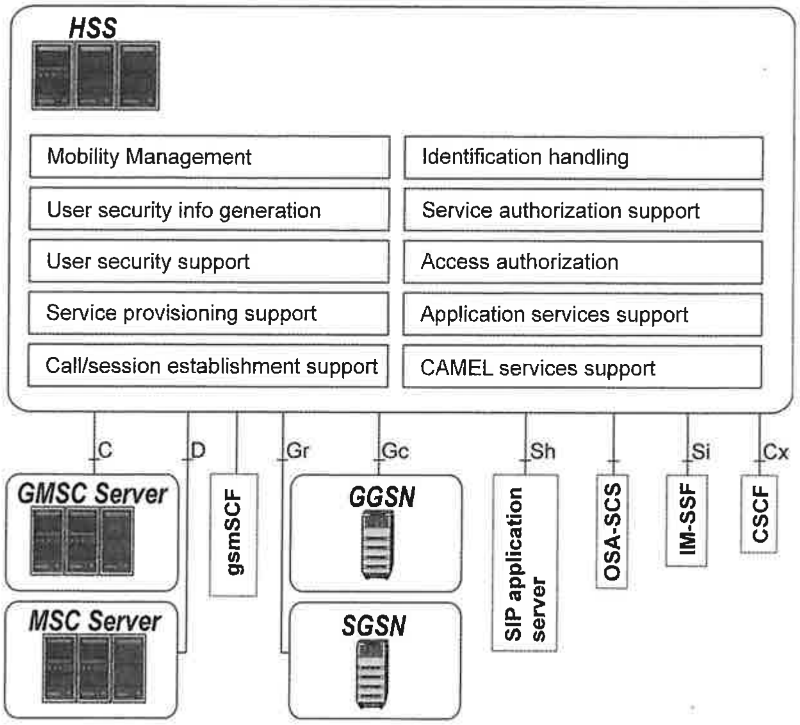


Figure 6.5 Logical diagram about Home Subscriber Server (HSS) functionalities and interfaces to Core Network (CN) domains

If we look at Figure 6.5, we see that the BC domain is not included. Although it has been defined to be part of the CN, its implementation with the 3G network is for further study.

We can see from Figure 6.5 that the majority of HSS functionalities are the ones that have existed in the network for a long time. They used to be taken care of by separate elements: the Home Location Register (HLR) and the Authentication Centre (AuC). In the 3GPP R5 architecture, the HLR and AuC are considered *HSS subsets*, but they still provide the same functionalities:

- Mobility Management (MM) functionality supports user mobility through the CS domain, PS domain and IMS. In this role the HSS, for instance, stores addressing information that can pinpoint the user/terminal location within the MM hierarchy.
- User security information generation, user security support and access authorization: these functionalities are mainly taken care of by the AuC subset, which sends signals to the CN domains and subsystems through the HLR subset.
- Service-provisioning support: the HSS provides access to the service profile data that are used within the CS domain, PS domain and/or IMS application services and Customised Applications for Mobile network Enhanced Logic (CAMEL) services support. The HSS communicates with the Session Initiation Protocol (SIP)

Application Server (AS) and the Open Service Architecture/Service Capability Server (OSA/SCS) to support application services in the IM CN subsystem. It also communicates with the IM-SSF to support CAMEL services related to the IM CN subsystem and with the GSM SCF to support CAMEL services in the CS domain and PS domain.

- Call/session establishment support: the HSS supports the call and/or session establishment procedures in the CS domain, PS domain and the IMS. For terminating traffic, it provides information about which call and/or session control entity is currently hosting the user.
- Identification handling: the HSS provides the appropriate relations among all the identifiers uniquely determining the user in the system: IMSI and MSISDNs for the CS domain; IMSI, MSISDNs and IP addresses for the PS domain; private identity and public identities for the IM CN subsystem. These items are discussed in more detail in Section 6.2.1.1.
- Service authorisation support: the HSS provides the basic authorisation for MT call/session establishment and service invocation. Furthermore, it updates the appropriate serving entities with the relevant information related to the services to be provided to the user.

In addition to the HSS, the Equipment Identity Register (EIR) is a functionality common to all domains and subsystems. The EIR stores information about end-user equipment and the status of this equipment. To do this, it makes use of three "lists": put roughly, the white list contains information about approved, normal terminal equipment; the black list stores information about stolen equipment; and the grey list contains serial number information about suspect equipment. Of these lists, the black and grey ones are normally implemented—it is unusual for the white list to be used. The EIR maintains these lists and provides information about user equipment to the CN Domain on request. If the EIR indicates that the terminal equipment is blacklisted, the CN domain refuses to deliver traffic to and from that terminal. In case the terminal equipment is on the grey list, the traffic will be delivered but some trace activity reporting may occur.

6.1.2 CS Domain

The CS domain is included in 3GPP R5 because the network must support CS services for backward compatibility. 3GPP R99 introduced CS domain structure that it directly inherited from the Global System for Mobile Communication (GSM). In 3GPP R4 the CS domain structure was given an alternative implementation method, where an operator had the ability to fine-tune CS domain control and traffic delivery capacity separately (Figure 6.6).

The aim of CS-MGW–MSC server division is to separate the control and user plane from each other within the CS domain. This introduces scalability to the system, since a single MSC server could control many CS-MGWs. Another advantage of this distributed CS domain architecture is that it opens up the possibilities for user plane geographical optimisation. For instance, an operator could locate CS-MGWs freely within its network and, by proper routing arrangements, it will be possible to arrange

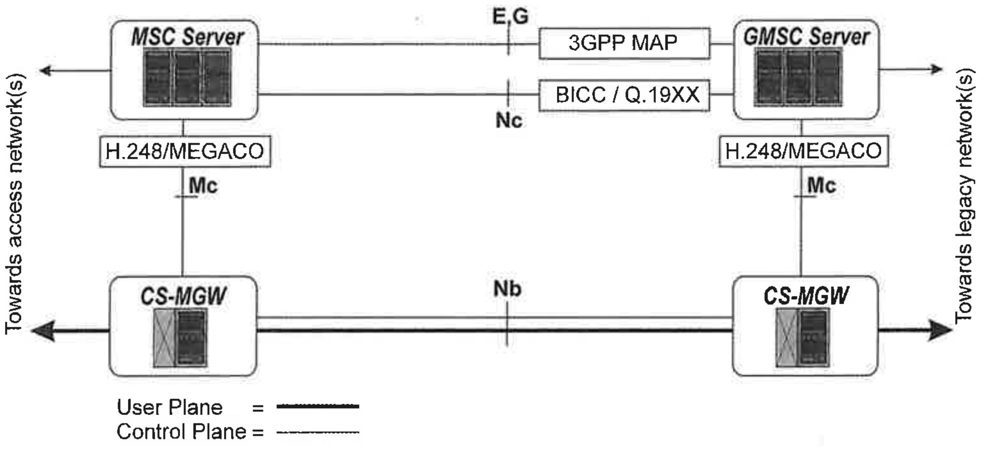


Figure 6.6 UMTS Core Network Circuit Switched (CN CS) domain with distributed Mobile Switching Centre (MSC) functionality (3GPP R4)

things in such a way that the user plane goes through the network geographically in the shortest possible way. The CS-MGW may also contain various conversion packages, which would give the operator the possibility of considering optimised transport network arrangements. For example, using the CS-MGW concept the operator could convert the CS domain backbone to use IP instead of other transport network mechanisms between the access network edge CS-MGW and the legacy Public Switched Telephone Network (PSTN) edge gateway.

The 3GPP-R4-distributed CS domain architecture defines MSC division, where call control functionality and the VLR are brought into an entity called the "MSC server". Respectively, user plane connectivity and related items (e.g., network inter-working) are brought into an entity called the "Media Gateway" (MGW). The CN as a whole contains all kinds of gateways and, thus, it is recommended to add the lettering "CS" in front of MGW to make it crystal clear that we are speaking about the Circuit Switched domain Media Gateway (CS-MGW).

When MSC server–CS-MGW separation is implemented, it opens up a new interface within the CS domain. That interface, Mc, uses the ITU-T H.248-defined Media Gateway Control Protocol (MGCP) for its purposes. H.248 only forms the basis for the information transfer mechanisms in this interface—the complete implementation contains various 3GPP-specific extensions. The Mc interface carries both *call-independent* and *call-dependent* H.248 transactions. The call-independent transactions in this interface contain mechanisms that control the way the CS-MGW imparts its functioning state to the MSC Server. Call-dependent transactions, in turn, can be visualised as "envelopes" transferring the control plane information that is either coming from a user through an access network or from a legacy network. Both call-independent and call-dependent transactions were first described in 3GPP TS29.232 version 5.0.0.

The Nc interface carries network–network call control information. In principle, any call control protocol is suitable for this purpose, as long as the protocol supports a call bearer and its control flow separation. For this purpose the 3GPP adopted a control protocol called "Bearer Independent Call Control" (BICC). To be exact, BICC is not a unique protocol; instead, it is a combination of various packages defined to be used together. These packages are mainly defined in ITU-T specification Q.1950 "Bearer independent call bearer control protocol".

The Nb interface carries both the user plane and the so-called transport network control plane. At the user plane end the Nb interface contains suitable frame protocols and other mechanisms for user data transfer. According to the relevant specifications, the Nb interface can be implemented either using ATM or IP transport. Both transport options were first covered in 3GPP TS29.414 version 5.0.0.

Of course, MSC servers need to communicate with each other. A couple of situations that could initiate this communication include MSC–MSC handover in GSM/EDGE Radio Access Network (GERAN) or the serving Radio Network Controller (RNC) relocation procedure occurring in UMTS Terrestrial Access Network (UTRAN), respectively. In these situations the control of user traffic moves from one MSC server to another and at the same time the CS-MGW on the edge of access network will change as well. These are the reasons MSC servers have Mobile Application Protocol (MAP) interfaces E and G, which transfer MM and other related information between MSC servers. For a detailed description of MAP see 3GPP TS29.002.

Maybe surprisingly, the 3GPP R5 CS domain does not need to be implemented according to 3GPP R4 directives. Another alternative is to continue with 3GPP R99 implementation at the CS end of the network. In this case the CS domain is directly inherited from the GSM world and follows GSM's traditional functionality. The advantage of this approach is that the need to invest in the network is smaller. However, there are also some drawbacks. By keeping the 3GPP R99 architecture within the CS domain the operator may lose the possibility of scalability. In addition, optimal user plane routing is not feasible with this solution, where the control and user plane are not so strictly separated from each other.

6.1.3 PS Domain

The two main elements of the PS domain are types of mobile network-specific servers: Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN).

SGSN contains the location registration function, which maintains data needed for originating and terminating packet data transfer. These data are subscription information containing the International Mobile Subscriber Identity (IMSI, see Section 6.2.1.1), various temporary identities, location information (see Section 6.2.1.2), Packet Data Protocol (PDP) addresses (de facto but not necessarily IP addresses), subscripted QoS (see Chapter 8) and so on.

The tool for data transfer within the PS domain is called the "PDP context" (see Section 6.2.2.2). In order to transfer data, the SGSN must know with which GGSN the active PDP context of a certain end-user exists. It is for this purpose that the SGSN stores the GGSN address for each active PDP context. Note that one SGSN may have active PDP contexts going through numerous GGSNs.

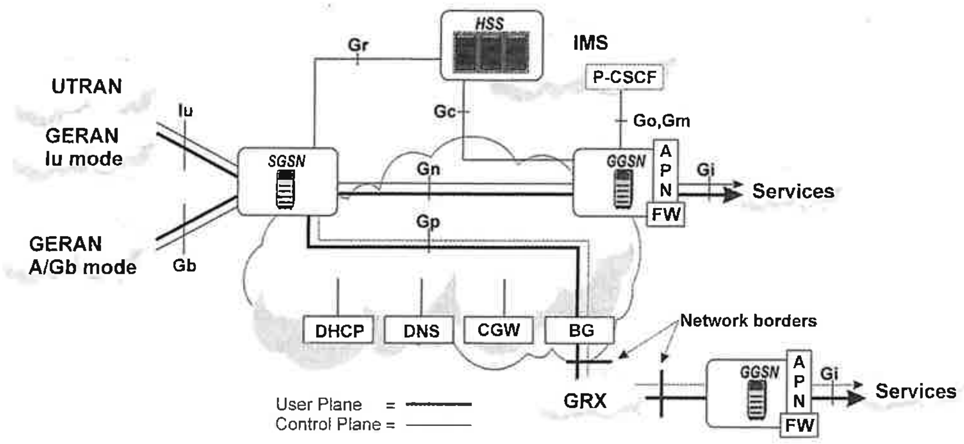


Figure 6.7 Rough guide to the Packet Switched (PS) domain structure

The GGSN also holds some data about the subscriber. These data may also contain the IMSI number, PDP addresses, location information and information about the SGSN that the subscriber has registered.

As far as the PS domain architecture is concerned, the SGSN and GGSN as such are insufficient. Packet traffic require additional elements/functionalities for addressing, security and charging. Figure 6.7 aims to illustrate the most relevant functionalities within the PS domain.

For security reasons operators now use dynamic address allocation for end-users. These addresses can be allocated in many ways, but the normal way to do this is to use Dynamic Host Configuration Protocol (DHCP) functionality/server. Depending on the operator's configuration, the DHCP allocates either IPv4 or IPv6 addresses for the end-user's terminal equipment.

Actually, the PS domain is, in a way, a sophisticated intranet. In order to address the various elements within this intranet, the Domain Name Server (DNS) is needed. The DNS within the PS domain is responsible for addressing PS domain elements. For example, when an SGSN establishes traffic to a certain GGSN, the SGSN requests the required GGSN address from the DNS.

When a user has gained a dynamically allocated address and the connection has been established between the SGSN and GGSN, the user is ready to access services that are made accessible by the operator. Service access is arranged through Access Point Names (APNs) which can be freely defined but very often are service-specific. For instance, one APN could be the "Internet" and through this APN the user is able to start Internet-browsing. Another APN could be, say, the "WAP" and this APN leads the end-user to browse WAP menus made available by the operator. One GGSN may contain tens of thousands of APN definitions: they could be company/corporate-specific, they could lead to any place, any network, etc. If the operator does not want to have this kind of access control, a so-called "wild card" APN can be brought into use. In this case end-user preferences as such are allowed and the operator just provides the connection.

Since security is an issue, the GGSN has a FireWall (FW) facility integrated. Every connection to and from the PS domain is done through the FW in order to guarantee security for end-user traffic.

There are many networks that contain a PS domain and roaming between these networks is a most vital issue as far as business is concerned. The PS domain contains a separate functionality in order to enable roaming and to make an interconnection between two PS domains belonging to separate networks. This functionality is called the "Border Gateway" (BG). GPRS Roaming Exchange (GRX) is a concept designed and implemented for General Packet Radio Service (GPRS) roaming purposes.

For charging data collection purposes the PS domain contains a separate functionality called the "Charging Gateway" (CGW). The CGW collects charging data from PS domain elements and relays them to the billing centre to be post-processed. Charging is also the main factor behind some GRX roaming arrangements. A very typical way of doing this is when a user is visiting a GPRS-capable network: the GGSN for GPRS connection is arranged from the home network of the user. By doing this the home network operator is in a position to collect charging data related to this GPRS connection. This arrangement also relinquishes control about APNs to the home network operator. Referring to the APN explanation above, this "home network GGSN" arrangement does not allow wild card APNs. If a visited network GGSN was used during roaming, wild card APNs are allowed, respectively.

As Figure 6.7 states, the PS domain maintains various connections. First, it maintains the IuPS interface towards access networks. Through this interface UTRAN and GERAN are connected. When GERAN is connected to the network in this way, it is said that the network uses *GERAN Iu mode*. There is still a possibility to use a frame-relay-based Gb interface for GERAN connection. In this case it is said that the network uses *GERAN Gb mode*. UTRAN is restricted to using the Iu interface for PS domain connections. Possible complementary accesses and their interconnection mechanisms are under study.

Second, the PS domain has a connection to CN common functionalities, like HSS and EIR. Through these connections the PS domain handles information related to the tasks presented in Section 6.1.1.

The PS domain is the network platform for sophisticated multimedia services enabled and maintained by the IMS. Thus, the PS domain contains interfaces towards the IMS. The IMS and its architecture are explained in Section 6.4.

6.2 CN Management Tasks and Control Duties

The previous section provided a short overview of the architectural aspects of the CN. In this section we will follow a slightly different approach: we will study the role of the CN through its management tasks and control duties.

As shown in Figure 6.8, as far as Communication Management (CM) is concerned, the two main tasks are connection management and session management. Connection management is the management task responsible for CS transactions and related issues, session management could be considered to be its counterpart at the PS end of the network. The control protocols carrying CM information, which deal with call and

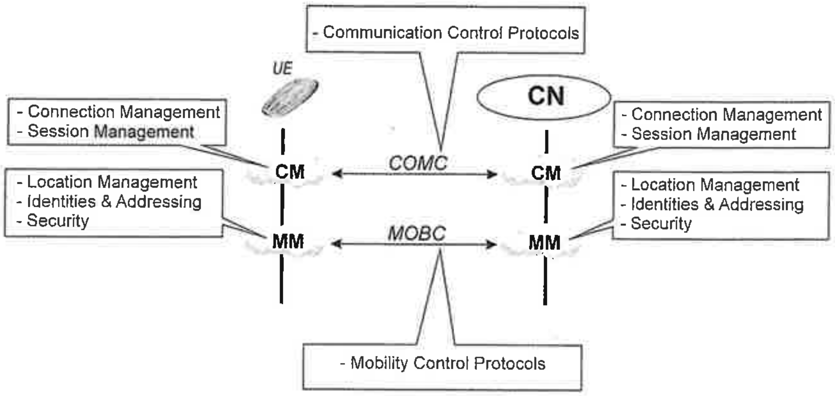


Figure 6.8 Core network (CN) management tasks and control duties

session control, are referred to here as a set of Communication Control (COMC) protocols.

The MM task covers the management of User Equipment (UE) locations together with their identities and addresses, related issues (security is also considered a part of MM). Security is discussed in Chapter 9 in more detail. The control protocols supporting execution of MM tasks are referred to as Mobility Control (MOBC) protocols.

6.2.1 Mobility Management (MM)

With worldwide 2G cellular network mobility is here to stay in communication networks. Understanding the essence of mobility makes the mobile network design significantly different—though more complex as well—from fixed communications and creates a lot of potential for provision of completely new kinds of services to end-users.

Let us first clarify the difference between two basic concepts related to user mobility:

- Location.
- Position.

The term "location" is used to refer to the location of the end-user (and his or her terminal) within the logical structure of the network. The identifiable elements within such a logical structure are the cells and areas (composed of groups of cells). Please note that the word "area" does not need to refer to a set of geographically neighbouring cells, but is simply used by the network operator for network operation purposes.

The term "position", on the other hand, refers to the geographical position of the end-user (and his or her terminal) within the coverage area of the network. The geographical position is given as a pair of standardised coordinates. In the most elementary case, when no geographical position can be determined, the position may be given as a cell identity, from which the position can be derived (e.g., as the geographical coordinates of the BS site controlling that cell).

Although both location and position deal with the whereabouts of this user, the answers are used by the UMTS network in a completely different manner. Location information is used by the network itself to reach end-users whenever there is a communication service activity addressed to them. Position information is determined by the UMTS network when requested by some external service (e.g., an emergency call centre). Although position information may well be life-critical to the end-user making the emergency call, location information is "life-critical" to the network itself in being able to provide services to mobile users in an uninterrupted manner.

Note that the primary purpose of positioning is to support application-oriented services—position information could also be utilised internally by the network. Examples of internal applications are position-aided handover and network-planning optimisation.

Mobile positioning as a service enabler is introduced among other services in Chapter 8.

Another key service created by the MM is *roaming*. The MM functions inside a single Public Land Mobile Network (PLMN) allowing a UMTS user to move freely within the coverage area of that single PLMN. Roaming is a capability that also enables users to move from one PLMN to another operated by a different operator company and, possibly, even in a different country. For the purpose of roaming many of the interfaces defined to be used in the CN are inter-operator interfaces, which are used by the CN elements in visited serving networks to retrieve subscriber location and subscription information from their home networks.

As stated, MM needs a kind of logical, relative hierarchy for its functioning. In addition to this structure, MM handles identities (permanent and temporary) and the addressing information of the subscribers and their terminals as well as those of the network elements involved. In the specifications, security aspects are also counted as part of MM.

6.2.1.1 Identities and Addressing of Users and Their Terminals

Unlike in fixed networks, the UMTS network requires the use of many kinds of numbers and identities for different purposes. In fixed networks the location of the subscriber and the equipment is, like the name says, fixed and this in turn makes many issues constant. When the location of the subscriber is not fixed, the fixed manner of numbering is no longer valid. The purposes of the different identities used in the UMTS can be summarised as follows:

- *Unique identity*: this is used to provide a globally unique identity for a subscriber. This value acts as a primary search key for all registers holding subscriber information; it is also used as a basis for charging purposes.
- *Service separation*: especially in the case of mobile terminating transactions the service going to be used must be recognised. This is done by using an identity having a relationship to the unique identity of the subscriber.
- *Routing purposes*: some special arrangements are required to perform transaction routing; this is not fixed to any network and country borders.

- *Security*: security is a very important topic in the cellular environment and this is why additional identities are generated to improve the privacy of users. Basically, these security-related identities are optional, but it is strongly recommended to use them anyway.

6.2.1.1.1 International Mobile Subscriber Identity (IMSI)

The unique identity for the mobile subscriber is called the *International Mobile Subscriber Identity* (IMSI). The IMSI consists of three parts:

$$\text{IMSI} = \text{MCC} + \text{MNC} + \text{MSN}$$

where MCC = the mobile country code (three digits), MNC = the mobile network code (2–3 digits) and MSN = the mobile subscriber number (9–10 digits). This number is stored in the SIM card (USIM). The IMSI acts as a unique database search key in the HLR, VLR, AuC and SGSN. This number follows the ITU-T specification E.214 on numbering. When the mobile user is roaming outside the home network, the visited serving network is able to recognise the home network by requesting the UE to provide this number. Because the IMSI is a unique database key for the subscriber's HLR located in the subscriber's home network, the HLR is able to return the subscriber profile and other information when requested by the IMSI. The same procedure is applied for security information requests from the home network.

6.2.1.1.2 Mobile Subscriber ISDN Number (MSISDN) and PDP Context Address

The IMSI number is used for exact subscriber identification. The *Mobile Subscriber ISDN Number* (MSISDN) is then used for service separation. Because one subscriber may have several services provisioned and activated, this number acts as a separator between them. For instance, the mobile user may have one MSISDN number for a speech service, another MSISDN number for a fax and so on. In the case of

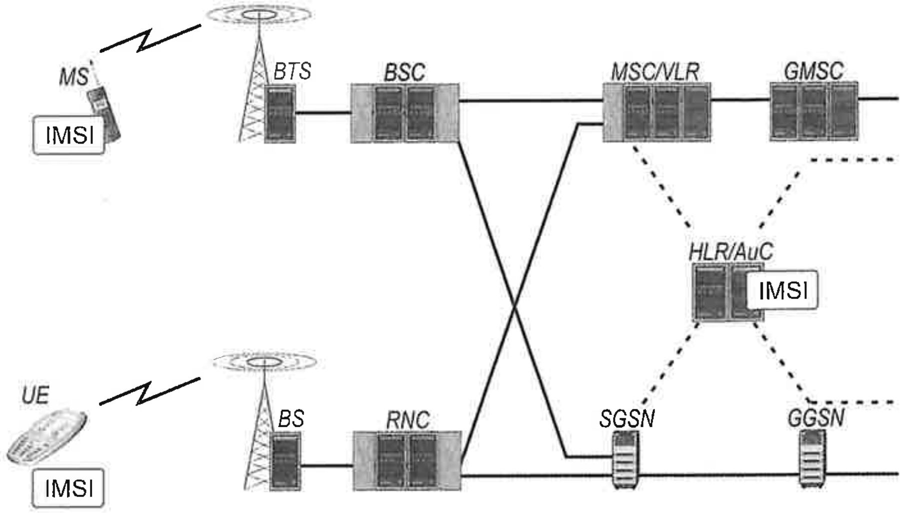


Figure 6.9 International Mobile Subscriber Identity (IMSI)

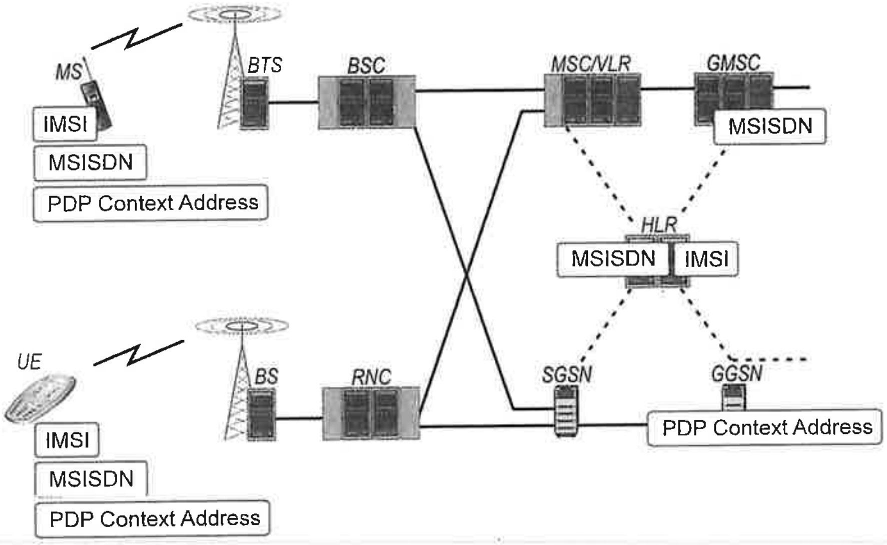


Figure 6.10 Mobile Station ISDN (MSISDN) number and Packet Data Prococol (PDP) context address

mobile-originated transactions the MSISDN is not required for service separation because indication of the service is provided within the CM message(s) during transaction establishment. In the mobile-terminated direction different MSISDN numbers are required for different services because the surrounding networks are not necessarily able to provide the service information by other means. The MSISDN consists of three parts:

$$\text{MSISDN} = \text{CC} + \text{NDC} + \text{SN}$$

where CC = the country code (1–3 digits), NDC = the national destination code (1–3 digits) and SN = the subscriber number. This number format follows the ITU-T specification E.164 on numbering. Very often this number is called a "directory number" or just simply a "subscriber number".

The functional PS counterpart for MSISDN is the *PDP context address*, which is an IP address of the mobile user. The PDP context address can be either dynamic or static. If it is dynamic, it is created when a packet session is created. If it is static, it has been defined in the HLR. If the PDP context address is static it behaves like an MSISDN at the CS end of the network.

6.2.1.1.3 Mobile Subscriber Roaming Number (MSRN) and Handover Number (HON)

The *Mobile Subscriber Roaming Number* (MSRN) is used for call-routing purposes. The format of the MSRN is the same as MSISDN (i.e., it consists of three parts, CC, NDC, SN and follows the E.164 numbering specification).

MSRN is used in mobile-terminated call path connection between the GMSC and serving MSC/VLR. This is possible because the allocated MSRN number recognises

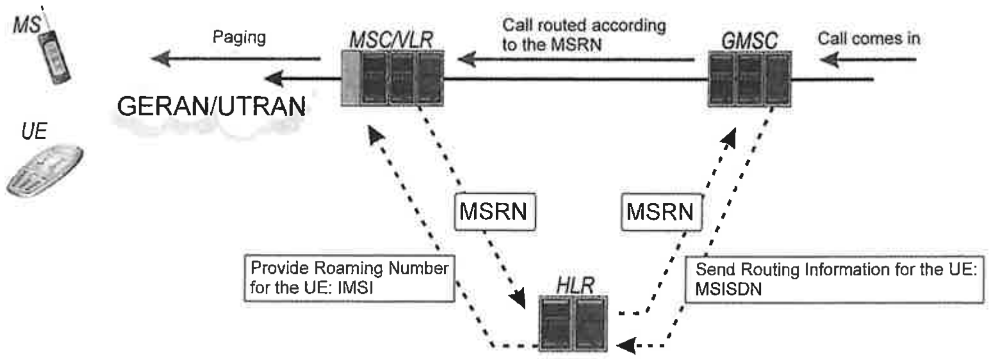

MS
UE
Paging
GERAN/UTRAN
MSC/VLR
Call routed according to the MSRN
GMSC
Call comes in
MSRN
MSRN
Provide Roaming Number for the UE: IMSI
HLR
Send Routing Information for the UE: MSISDN

Figure 6.11 Mobile Subscriber Roaming Number (MSRN)

the country, network and the network element within the network. The "subscriber part" of MSRN is for subscriber recognition. The MSRN is also used for call path connection between two MSC/VLRs in the case of MSC–MSC handover. In this context the MSRN is often called the *Handover Number* (HON).

6.2.1.1.4 Temporary Mobile Subscriber Identity (TMSI) and P-TMSI

For security reasons it is very important that the unique identity IMSI is transferred to non ciphered mode as seldom as possible. For this purpose, the UMTS system makes use of the *Temporary Mobile Subscriber Identity* (TMSI) instead of the original IMSI. The PS domain of the CN allocates similar temporary identities for the same purpose. In order to separate these from the TMSI, they are called *Packet Temporary Mobile Subscriber Identities* (P-TMSIs).

The TMSI and P-TMSI are random format numbers, which have limited validity time and validity area. TMSI numbers are allocated by the VLR and are valid until the UE performs the next transaction. The P-TMSI is allocated by the SGSN and is valid over the SGSN area. The P-TMSI changes when the UE carries out a routing area update.

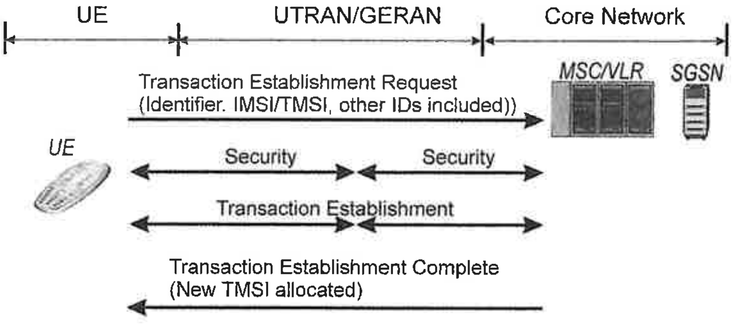

UE
UTRAN/GERAN
Core Network
Transaction Establishment Request
(Identifier. IMSI/TMSI, other IDs included))
MSC/VLR
SGSN
UE
Security
Security
Transaction Establishment
Transaction Establishment Complete
(New TMSI allocated)

Figure 6.12 Allocation of Temporary Mobile Subscriber Identity (TMSI)

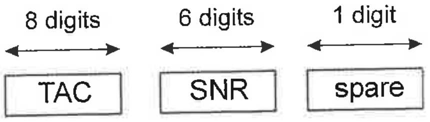


Figure 6.13 International Mobile Equipment Identity (IMEI)

6.2.1.1.5 International Mobile Equipment Identity

Two slightly different numbers are defined for mobile equipment identification purposes: the International Mobile Equipment Identity (IMEI) and its extension, the International Mobile Equipment Identity and Software Version (IMEISV). Both of these numbers are handled by the EIR. The network procedures are similar for both numbers: the UE provides either of these numbers upon request and the network verifies the status of the number with the EIR. The structures of the number are as described in Figures 6.13 and 6.14.

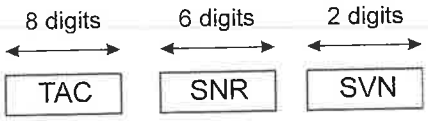


Figure 6.14 International Mobile Equipment Identity and Software Version (IMEISV) number

Both of these numbers have common parts: the Type Allocation Code (TAC) defines the manufacturer and type of phone; and the Serial NumbeR (SNR) uniquely defines a piece of equipment within the TAC. The defined length of the IMEI is 15 digits and the remaining digit is spare, which the UE fills with the value "0" when sending it to the network. The network takes the 14 most significant digits of the IMEI and computes the check digit value in order to verify transmission correctness.

The IMEISV is two digits longer as a result of the IMEI containing the software version number of the hardware expressed by 2 digits. Either usage of these numbers is allowed: the UE may send the IMEI or IMEISV but not both.

It is also worth noting that the IMEI remains unchanged as it is the unique identity of the piece of hardware. In the case of IMEISV, the SVN part of the number will change in the context of software updates but the rest of the number remains unchanged.

6.2.1.1.6 IMS—Home Network Domain Name

The IMS home network domain name follows the naming structure as defined for Internet use but contains mobile-network-specific parts. To make a separation between a mobile network and an Internet namespace, the IMS home network domain name contains parts of an IMSI number (see Section 6.2.1.1.1).

When, for instance, the UE wants to ascertain the home network domain name, the procedure involves the following steps:

1. The MNC and MCC are ascertained from the IMSI number.
2. The home network domain name always starts with the label "ims".

3. The home network domain name always ends with the label "3gppnetwork.org".
4. The home network domain name is formed by joining parts (2) and (3) together as, for instance, "ims.mnc<MNC>.mcc<MCC>.3gppnetwork.org".

As an example, suppose we have an IMSI number 244 182 123123123. In this IMSI, MCC = 244, MNC = 182 and the Mobile Subscriber Identification Number (MSIN) = 123123123. Using this example IMSI, we can resolve the IMS home network domain name as:

ims.mnc182.mcc244.3gppnetwork.org

As explained in Section 6.2.1.1.1, in UMTS-based networks the MNC comprises 3 digits. However, GSM-based networks mostly use 2-digit MNC values. If the MNC of the network is 2 digits, an additional "0' is added in front of the MNC in the home network domain name.

6.2.1.1.7 IMS—Private User Identity

The IMS requires private user identities for its functionality. These follow the Internet naming structure: *username@realm*. A private username could in principle be anything, but in practice UMTS usernames are derived from an IMSI number, since an IMSI constitutes a good means of providing a unique and confidential identity for the user.

If we use the same IMSI number as in the previous example, the private user identity will be as follows:

IMSI	244 182 123123123
MCC	244
MNC	182
MSIN	123123123
Realm	ims.mnc182.mcc244.3ppnetwork.org
Username	IMSI

Using these values, the private user identity in this example will be:

244182123123123@ims.mnc182.mcc244.3gppnetwork.org

A private user identity looks very much like an IMSI and behaves in part similarly. It is stored in an IMS Identity Module (ISIM) and has the following main characteristics:

- It is contained in all registration requests passed from the UE to the network (IMS).
- The IMS holds the registration and deregistration status of private user identities.
- It is permanently allocated and valid as long as subscriptions are maintained.
- The UE is not able to modify its private user identity under any circumstances.
- The HSS stores private user identities.

6.2.1.1.8 IMS—Public User Identity

A private user identity can also be used for purposes that are internal to the network. For reachability and addressing reasons a user must clearly have a public user identity. Since we are talking about a mobile user, he or she must be reachable in two ways: through the Internet using the Internet's addressing style or through conventional mobile equipment using an E.164 MSISDN type of address (see Section 6.2.1.1.2).

An Internet-style public user identity is actually a SIP URI. For example, its format looks like:

sip:firstname.lastname@operator.com

A telephone-style public user identity follows the standard MSISDN number definition as stated above. Suppose that the MSISDN number is +358 66 1231234. When it is expressed as a Tel URL it looks like:

tel: + 358661231234

It is also possible to indicate what kind of numbering plan the Tel URL follows. The example shown above expresses telephone numbers according to the global numbering plan. If a local numbering plan is used (country code with the network code possibly left out) the scope and owner of the numbering plan must be indicated within the Tel URL.

A public user identity has the following main characteristics:

- One of two possible formats: SIP URI or Tel URL.
- At least one public user identity is stored in the ISIM.
- It must be registered before it can be used with IMS sessions and services.

One user could have many public user identities.

6.2.1.2 Location Structures and Their Identities

In addition to the addressing and identities of the subscribers and their terminals the MM requires the network to have a logical structure. This structure can be represented as logical parts of the access network. Thus, these logical entities act like a "map" for MM procedures and their parameterisation. The UMTS basically contains four logical definitions:

- Location Area (LA).
- Routing Area (RA).
- UTRAN Registration Area (URA).
- Cell.

In the CN CS domain the LA is the area where the UE can freely move without performing a location update procedure. The LA consists of cells: the minimum is one cell and the maximum is all the cells under one VLR. In a location update procedure the location of the UE is updated in the VLR with LA accuracy. This

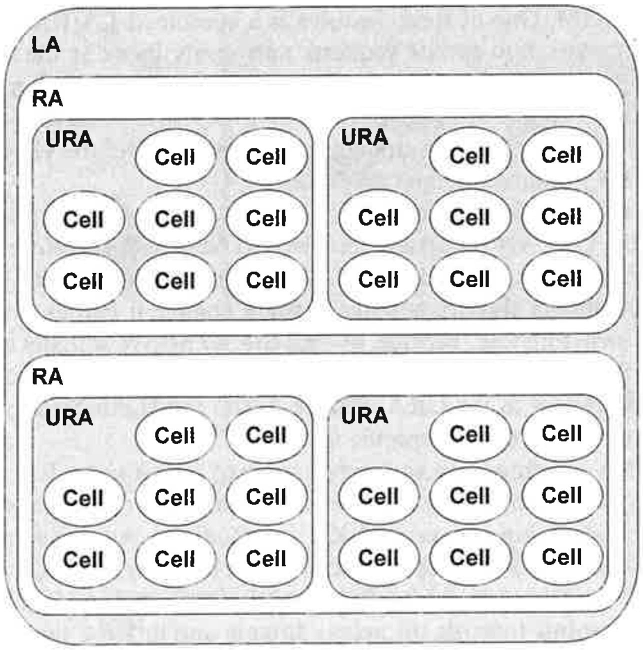


Figure 6.15 Mobility Management (MM) logical entities and their relationships

information is needed for mobile-terminated calls; to get this information the VLR pages the desired UE in the LA where the UE last performed a location update.

Note that in all other respects (other than the VLR) the LA does not have any hardware constraints. For instance, one RNC may have several LAs or one LA may cover several RNCs. Every LA is uniquely identified with a *Location Area Identity* (LAI). The LAI consists of the following parts:

$$\text{LAI} = \text{MCC} + \text{MNC} + \text{LA code}$$

where the MCC and MNC have the same format as in the IMSI number. The LA code is just a number identifying an LA. The LAI is a globally unique number, and within the same network the same LA code should clearly not be repeated as a single VLR cannot handle duplicate LA codes. The UE listens to the LAI(s) from the Broadcast Channel (BCH). The content of this transport channel is cell-specific and is filled by the RNC.

Like the CN CS domain, the PS domain has its own location registration procedure based on an RA. An RA is very similar to an LA (i.e., it is the area where the UE may move without performing an RA update). On the other hand, an RA is a kind of "subset" of LA: one LA may have several RAs within it but not vice versa. In addition, one RA cannot belong to two LAs.

The CS and PS domains may have an optional Gs interface between them (VLR and SGSN) through which these nodes may exchange location information. Since the UMTS must interoperate with the GSM, the UMTS CN also supports features available in the GSM. One of these features is a combined LA/RA update, where the GSM terminal carries out update requests and sends them in the first place to the SGSN. If an optional Gs interface is available, the SGSN also uses this interface to request the VLR to update LA registration. In a normal UMTS network (i.e., without this option) the combined LA/RA update is not available and the UE has to register its location to both CN domains separately.

In the GSM network, MM is completely handled between the terminal and the NSS. In the UMTS the UTRAN is partially involved in MM and, therefore, contains a local mobility registration: this is called a *UTRAN Registration Area* (URA) and is discussed in Chapter 5. Although this sounds like a small change it causes marked changes in SGSN internal structure and, because of this, the 3G SGSN actually contains both 2G SGSN and 3G SGSN functionality. In the UMTS the SGSN carries tunnelled IP traffic to/from a UE according to the URA identity. In 2G the SGSN terminates tunnelled IP traffic and relays it over the 2G-specific Gb link.

Because a URA is defined in a very similar way to an LA or an RA, it does not have, in principle, any limitations insofar as network elements are concerned. In practice, it seems that the relationship between a URA and Radio Network Subsystems (RNSs) is more or less fixed. On the other hand, the URA is, in a way, a logical definition, which combines traffic routing and Radio Resource Control (RRC). In routing, the URA addressing entity points towards the access domain and in RRC the terminal has states indicating location accuracy and traffic reception ability. This is visible in the RRC-state model that was briefly presented in Chapter 5.

The smallest "building block" used for these MM logical entities is the "cell". Basically, the CN needs to be aware of sets of cells (i.e., areas) rather than individual cells. The cell in the access domain is the smallest entity that has its own publicly visible identity, called the *Cell ID* (CI). Like the LA code the CI is just a number that should be unique within the network. To globally separate cells from one another, the identity must be expanded, and in this case it is called a *Cell Global Identity* (CGI). The CGI has the following format:

$$\text{CGI} = \text{MNC} + \text{MCC} + \text{LA code} + \text{CI}$$

The CGI value covers the country of the network (MCC), the network within a country (MNC), the LA in the network and finally the cell number within the network. This information is distributed to the UE by the UTRAN functionality for system information broadcasting.

6.2.1.3 Network-level Identifiers Common to the CN and Access Network(s)

This section briefly introduces some identifiers that are transferred across the Iu interface and, thus, are common to the CN and the UTRAN. Some of these identities are also used when interconnecting with other networks.

As shown in Figure 6.16 every PLMN has its own, unique ID value called the PLMN-id. This value consists of two parameters: the MCC and the MNC. These are equivalent to the same parameters used in the IMSI.

$$\text{PLMN-id} = \text{MCC} + \text{MNC}$$

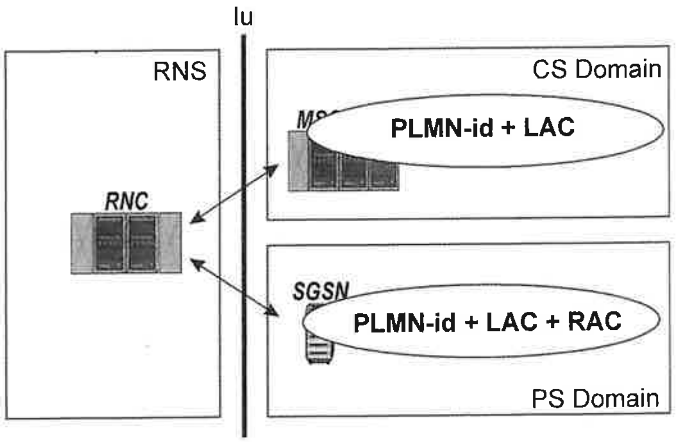


Figure 6.16 Core Network (CN) domain identifier

A PLMN-id is used for many purposes, since it is a very handy way to create a globally unique value that can be transferred between networks. For example, both the CGI and the LAI start with a PLMN-id.

Within a network the RNS must be able to locate the CN domain edge node that maintains the Iu interface. A CN domain identifier is used for this purpose. CN edge domain information is needed when a connection is established over the Iu interface and when a Serving RNS (SRNS) functionality is relocated (i.e., the bearers allocated for connection(s) over the Iu interface are changed so that another RNS can be used).

A CN CS domain identifier consists of a PLMN-id and a Location Area Code (LAC), while a CN PS domain identifier contains a PLMN-id, an LAC and a Routing Area Code (RAC).

Within a CN there may be a need to identify one of its elements globally. This can be done using a CN identifier (CN-id). This also sets theoretical limits for the number of CN element per network. A CN-id consists of two values: a PLMN-id and a numerical integer value between 0 and 4,095:

$$\text{Global CN-id} = \text{PLMN-id} + \langle 0 \ldots 4{,}095 \rangle$$

Note that this identifier is *not* the same as an element address. A CN-id is a globally unique ordinal number and, in addition to this, CN elements may or may not have MSISDN-type addresses used for Signalling Connection Control Part (SCCP) routing purposes.

An RNC identifier (RNC-id) follows exactly the same format as a CN-id and is used in RNC elements and in Base Station Controllers (BSCs) when GERAN is the network employed in Iu mode:

$$\text{Global RNC-id} = \text{PLMN-id} + \langle 0 \ldots 4{,}095 \rangle$$

A Service Area Identifier (SAI) identifies an area consisting of one or more cells belonging to the same LA and is recognised by all CN domains. It can be used to indicate the location of a UE to the CN domain. Figure 6.17 shows a sample SAI.

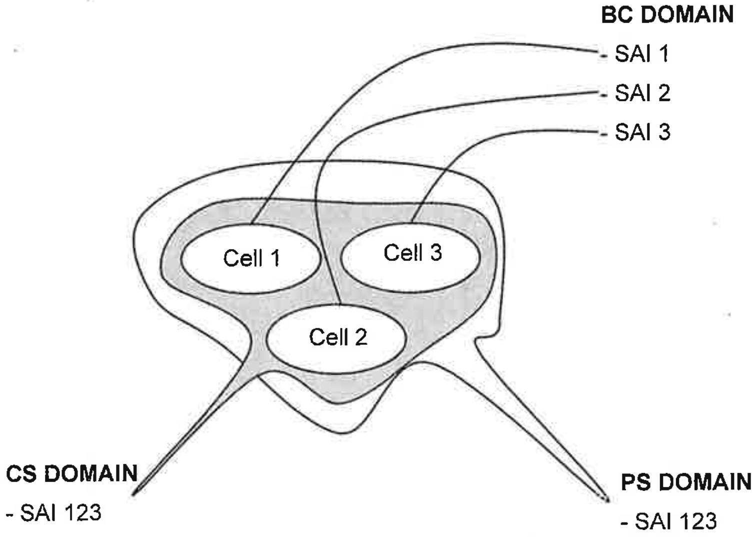


Figure 6.17 Service Area Identifier (SAI)

By studying Figure 6.17 we can see that the following characteristics and limitations apply:

- In both PS and CS domains a service area may consist of more than one cell. In the BC domain one service area is always one cell.
- A cell can have a maximum of two SAIs defined. In this case one is used for CS and PS domains and the other for the BC domain.
- The format of SAI is PLMN-id + LAC + SAC, where SAC stands for Service Area Code.

Due to licensing and implementation costs a commercial need for network sharing has been under discussion in many countries. Actually, network sharing is one of the topics in 3GPP R6 specification work currently under review. A network can be shared in many ways, but if sharing occurs there must be a way to identify which part of the network is shared. So far, a Shared Network Area Identifier (SNAI) has been introduced for this purpose.

A Shared Network Area (SNA) consists of LAs that are redefined to be an SNA. It is through this area that UEs gain access to different operator networks. An SNAI, or SNA-id, contains a PLMN-id and Shared Network Area Code (SNAC):

$$\text{SNA-id} = \text{PLMN-id} + \text{SNAC}$$

Since a PLMN-id is included in an SNA-id, the SNA-id is globally unique in nature.

6.2.1.4 Mobility Management State Model

The presence of a packet connection and its management brings in a new dimension as far as MM is concerned; for packet connections the MM has a state model. In CS

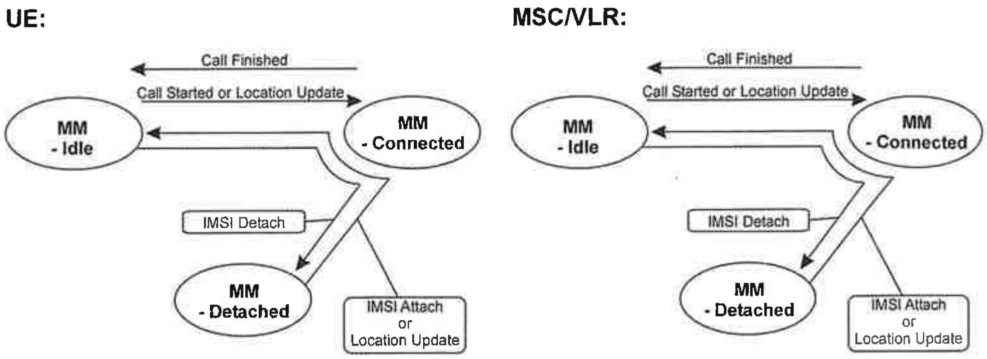


Figure 6.18 Circuit Switched (CS) Mobility Management (MM) state model

connections basically the same kind of model exists but it is rarely used because CS connection behaviour does not require this kind of state model. As already indicated in Figure 6.18, the abbreviation MM stands for CS MM, while PMM refers to PS MM.

6.2.1.4.1 MM States in CS Mode

From the MM point of view, a terminal's network connection may have three states: MM-detached, MM-idle and MM-connected. These MM states indicate how accurately the terminal location is known when compared with the logical structure presented in Figure 6.15. In the MM-detached state the network is not aware of the terminal/subscriber at all (i.e., the MM state when the terminal is switched off). In the MM-idle state the network knows the location of the terminal/subscriber down to the accuracy of an LA. In the MM-connected state the network knows the location of the terminal down to the accuracy of a cell.

The situation described in Figure 6.18 is similar to both the GSM Network Subsystem (NSS) and the UMTS CS domain (after 3GPP R99). When a subscriber switches his or her terminal on, the terminal performs either an IMSI attach or location update procedure. Accordingly, when the subscriber switches his or her terminal off the MM state changes from MM-idle to MM-detached. IMSI attach is performed if the LAI recognised by the camped cell is the same as the one the terminal has in its UMTS Service Identity Module (USIM). If the LAI received from the network differs from the one stored, the terminal performs location update to update and possibly register its new location within the CN CS domain and HLR. In either case the MM state changes as follows: MM-detached → MM-connected → MM-idle. The point we are trying to make is that when the terminal performs either of these briefly described procedures, the network is momentarily aware of the terminal location down to the accuracy of a cell. Both of these procedures cause the CS domain to "wake up" by delivering the same type of message containing originating cell information and the reason for the transaction. In the GSM this message is called a "CM service request" and in the UMTS it is called a "UE initial message".

When a subscriber is active (MM-idle, terminal switched on) the MM state is toggled between MM-idle and MM-connected states according to the use of the terminal. To put it simply: when a call starts the MM state goes from MM-idle → MM-connected and when the call is finished the MM state goes from MM-connected → MM-idle.

6.2.1.4.2 Mobility Management States in Packet Switched Mode

For PS connections the situation is different in terms of PMM procedures: PMM (Packet Mobility Management) states are the same but the triggers that allow movement from one state to another are different (Figure 6.19).

When the PMM state is PMM-detached the network does not have any valid routing information available for PS connection. If the PMM state needs to be changed an option is to perform packet IMSI attach. This procedure takes place whenever a terminal supporting PS operation mode is switched on. A confusing issue here is that this so-called packet attach is a very different procedure from its counterpart at the CS end of the network. Packet IMSI attach as a procedure is a relatively heavy signalling procedure reminiscent of location update. With packet IMSI attach the valid routing information for PS connection is "created" in every node involving PS connections: SGSN and GGSN. In addition to this, the subscriber profile is requested from the HLR and, possibly, old routing information is eliminated.

In PMM-connected state, data can be transferred between the terminal and the network: the SGSN knows the valid routing information for packet transfer down to the accuracy of the routing address of the actual SRNC. In PMM-idle state the location is known down to the accuracy of an RA identity. In PMM-idle state a paging procedure is needed to reach the terminal (e.g., for signalling).

From the end-user point of view, a PS mobile connection is often described as "being always on"; on the other hand, a packet call is said to be like many short CS calls. Both of these statements contain some truth but they are not quite exact as such. From the network point of view, a PS mobile connection gives the illusion of "being always on". This illusion is created by the PMM-connected and PMM-idle states. In PMM-idle state both the network and the terminal hold valid routing information and they are ready for packet data transfer but they are *not* able to transfer packets in this state since there is no connection present through the access network.

When the subscriber switches his or her packet-transfer-capable terminal off, MM reverts to the PMM-detached state and the routing information possibly present in the network nodes is no longer valid. If, for one reason or another, errors occur in the

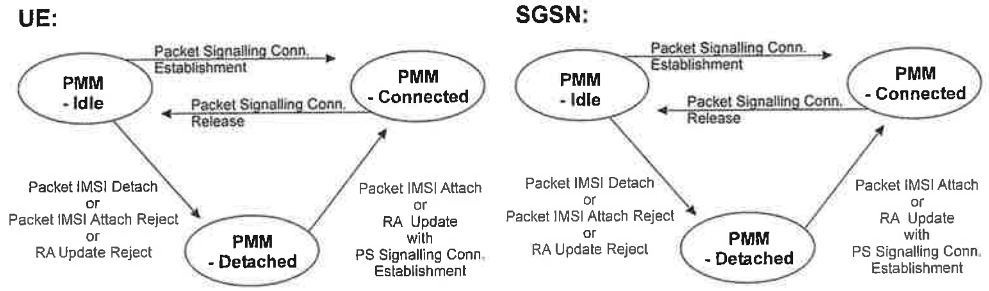

UE:
PMM - Idle
Packet Signalling Conn. Establishment
PMM - Connected
Packet Signalling Conn. Release
Packet IMSI Detach or Packet IMSI Attach Reject or RA Update Reject
PMM - Detached
Packet IMSI Attach or RA Update with PS Signalling Conn. Establishment
SGSN:
PMM - Idle
Packet Signalling Conn. Establishment
PMM - Connected
Packet Signalling Conn. Release
Packet IMSI Detach or Packet IMSI Attach Reject or RA Update Reject
PMM - Detached
Packet IMSI Attach or RA Update with PS Signalling Conn. Establishment

Figure 6.19 Packet Switched (PS) Mobility Management (MM) state model

context of packet IMSI attach or RA update, the MM state may return to PMM-detached.

6.2.2 Communication Management (CM)

In this subsection we briefly describe the main functions of CM in terms of both CS and PS communications. From a CS standpoint, the functionality is referred to as CM. From a PS standpoint, the entire functionality is referred to as Session Management (SM). These functionalities are covered by presenting the main phases of a connection or SM process as well as the entities.

6.2.2.1 Connection Management for Circuit Switching

Connection management is a high-level name describing the functions required for incoming and outgoing transaction handling within a switch. Generally speaking, the switch should perform three activities before a CS transaction can be connected. These activities are number analysis, routing and charging. Connection management can functionally be divided into three phases, which the transaction attempt must pass in order to perform through-connection (Figure 6.20).

Number analysis is a collection of rules on how the incoming transaction should be handled. The number of the subscriber who initiated the transaction is called the "calling number" and the number to which the transaction should be connected is the "called number". Number analysis investigates both of these numbers and makes decisions based on the rules defined. Number analysis is performed in both phases of connection management. In Phase I the switch checks whether the called number is obtainable and whether any restriction, such as call-barring, is to be applied to the calling number.

In Phase II the system concentrates on the called number. The nature of the transaction is investigated to ascertain whether it is an international or national call and whether there is any routing rule defined for the called number? In addition, the system checks whether the transaction requires any inter-working equipment (like a modem) to be connected and whether the transaction is chargeable or not. Also, the statistics for this transaction are initiated in this phase.

As a successful result of connection management Phase II, the system knows where the transaction attempt should be connected. This connection and channel selection procedure is called "routing". When the correct destination for the transaction is known the system starts to set up channel(s)/bandwidth towards the desired destination by using, for instance, ISDN User Part (ISUP) signalling protocol. During the transaction the switch stores statistical information about the transaction and its connection and collects charging information (if the transaction was judged to be chargeable). When the transaction is finished, connection management Phase III takes care of releasing all the resources related to the transaction.

In fixed networks every call is treated as an entity at both ends. In cellular networks the term "call" can be interpreted in many ways. Every "call" consists of call legs and each leg thus defines a part of a call.

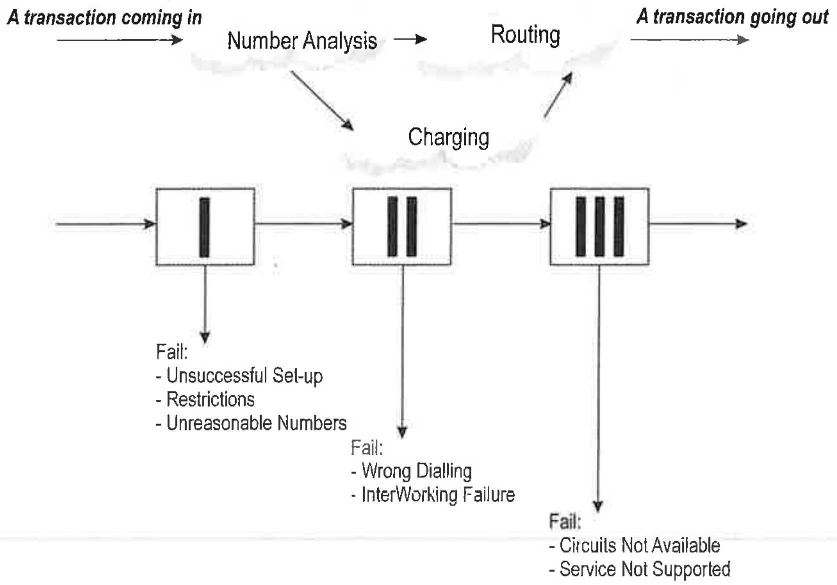


Figure 6.20 Circuit Switched (CS) connection management—connection diagram

From the connection management point of view, every call consists of at least two legs. There are four legs available: MOC (Mobile Originated Call), MTC (Mobile Terminated Call), POC (PSTN Originated Call) and PTC (PSTN Terminated Call). As Figure 6.21 indicates, connection management is actually a distributed functionality and, depending on which element is in question, different parts of call control are used. If a serving MSC/VLR is in question, it handles MOC and MTC call legs; and if the GMSC is in question, it handles POC and PTC legs. Call control is able to receive and create these call legs and also to determine whether any additional functionality is required based on the leg and call type. The most important additional facilities are network inter-working and charging.

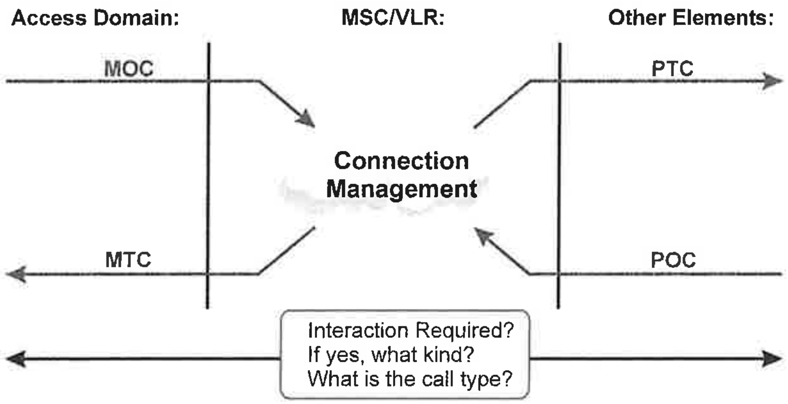


Figure 6.21 Call legs

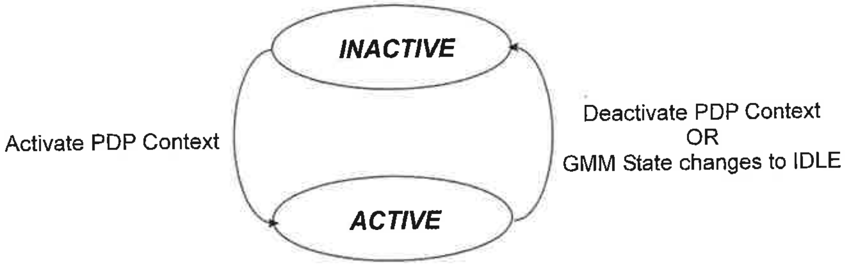


Figure 6.22 Session Management (SM) state model

Call control recognises the call type and based on this it decides further actions. The basic CS call types are:

- Normal call (voice).
- Emergency call.
- Data call (fax included).

In the MOC leg, this call type is included in the "CM service request message" (GSM) and "UE initial message" (UMTS with UTRAN). In the POC leg, the call type is "hidden" in the called party address (B Number); as explained in the context of MM, the service to be used is recognised in the terminating direction by the MSISDN number. In both MTC and PTC legs, connection management determines whether any kind of interaction is required between the call legs.

6.2.2.2 Session Management (SM) for Packet Communication

In the PS domain, packet connections are called "sessions" and they are established and managed by an entity called Session Management (SM).

SM as a logical entity has two main states: *inactive* and *active*. In the inactive state, packet data transfer is not possible at all and also routing information (if it exists) is not valid. In the active state, packet data transfer is possible and all valid routing information is present and defined.

The protocol used for packet data transfer during an active session is PDP. The design of the CN PS domain allows many different PDP protocols to be used. The most obvious case is to use IP as a PDP, but other protocols like X.25 could be supported, though these cases will rarely occur.

The SM handles packet session attributes as contexts and the term used here is *PDP context*. The PDP context contains all parameters describing packet data connection by means of end point addresses and QoS. For example, a PDP context holds such information as allocated IP addresses, connection type and related network element addresses. When SM is active (i.e., a PDP content exists), the user also has an IP address. From the service point of view, one PDP context is set up for one PS service with a certain QoS class. Thus, for instance, Web-surfing and streaming of video over packet connection have their own PDP contexts.

As defined, the UMTS uses the following QoS classes:

- Conversational class.
- Streaming class.
- Interactive class.
- Background class.

These classes are discussed in more detail in Chapter 8.

The PDP context is defined in the UE and the GGSN; and when it exists it contains all relevant parameters defining characteristics for packet connection, as described earlier. The PDP context can be activated, deactivated or modified.

Activation of the PDP context causes the SM to change its state from inactive to active. This SM state change in turn means that the UE forming a packet session with the network holds valid, allocated address information, and that the characteristics of packet connection are defined (e.g., QoS to be used). When the PDP context has been activated the UE and the network are able to establish a bearer for data transfer.

Deactivation of the PDP context causes the SM to change its state to inactive. When the SM state is inactive, the address information and packet session information the UE and the network may have is no longer valid. Thus, the UE and the network are not in a position to arrange any connections to transfer user data flows.

When the SM state is active and a PDP context exists, the PDP context can be modified. In this modification process the UE and the network renegotiate the packet session characteristics. A typical topic for this renegotiation is the QoS class of the packet session.

As shown in Figure 6.23, SM is a high-level entity and its activity depends on lower level entities: PMM and RRC. If RRC and PMM states are not suitable for the active packet session, the PDP context is deactivated and the SM state is changed from active to inactive. This kind of situation occurs, for example, when the RRC changes its state from connected to idle. This state change triggers the PMM to change its state from connected to idle and, as far as the SM state model in Figure 6.22 is concerned, the PMM state change triggers the SM to change its state from active to inactive.

The purpose of SM is to create the illusion of continuity in the connection to the end user, and this must be done in an effective way by saving network resources whenever possible; for instance, the Radio Access Bearers (RABs) carrying user data are established when required, and when there is nothing to transfer the RABs are cleared but the packet signalling connection still remains. Figure 5.19 shows how the different management and controlling entities within the CN and UTRAN change their states during an example packet data flow. This example describes a situation in which the subscriber turns his or her terminal on and IMSI attach is performed. After IMSI attach the subscriber sends some packet data to the network and the service provider sends some data packets to the terminal. This procedure is repeated and then, after some time, the network sends some packet data to the UE. Finally, the terminal is switched off. This example data flow can be realised, for example, during a WAP browsing session.

While the terminal is switched off there are no activities going on between the terminal and network. When it is switched on, an IMSI attach procedure is executed and the UE becomes identified by the network. The establishment of a signalling

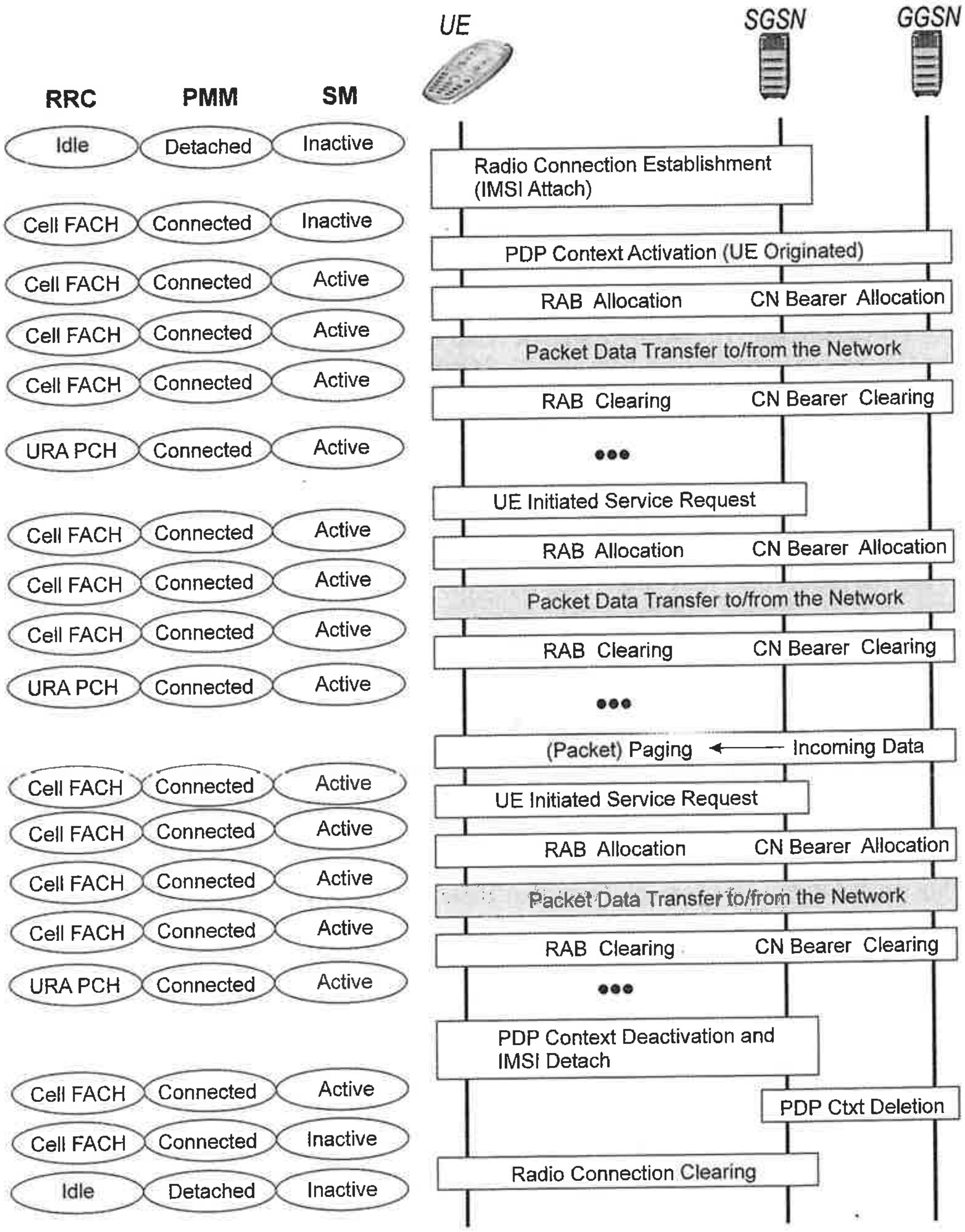


Figure 6.23 Radio Resource Control (RRC), Packet Mobile Management (PMM) and Session Management (SM) states during a packet flow (example)

connection brings the RRC state from "idle" to "connected cell (FACH)" (where FACH is the Forward Access Channel). At the same time the PMM state changes from "detached" to "connected" and the network now has valid information about the location of the subscriber. The SGSN gets this information in a "UE initial message" containing information about the desired activity. When the IMSI has been attached the UE initiates "PDP context activation". During this procedure the UE and the network negotiate the desired packet connection characteristics (e.g., the QoS class). The result of the "PDP context activation" procedure is that the SM state changes from inactive to active.

If there are any packet data to be transferred a bearer capable of carrying user data flows is established. The SGSN starts the RAB allocation procedure over UTRAN and the CN bearer is established between the SGSN and the GGSN. The network is now able to transfer packet data to and from the UE. The RRC state is used to optimise UTRAN resources; when the RRC is in the "connected cell FACH" state, only a small amount of packet data can be transferred over the Iu interface. In this RRC state the UE does not have a dedicated connection with the network. If the amount of packet data amount was larger, a dedicated channel would be allocated to the UE and the RRC state would be "connected cell DCH" (where DCH is the Dedicated Channel). When the packet data have been transferred the RAB and CN bearers carrying the user data flow are cleared but the PDP context is preserved. In addition, the signalling bearers forming the signalling connection between the UE and the network are maintained in this example. When the data bearers are cleared the RRC connection state changes to "connected URA PCH" (where PCH is the Paging Channel) to save UE resources. In this RRC state the network does not know the exact location of the UE, and if the network desires to communicate with the UE the UE must be paged.

If after a while the UE again desires to send packet data to the network it sends a "service request" message to the network and the state of the RRC connection changes once again to "connected cell FACH". The "service request" triggers the network to allocate RAB and CN bearers. Note that these bearers are allocated according to the parameterisation included in the PDP context. When these bearers have been established the packet data is transferred to and from the UE. Upon completion of packet data transfer the bearers carrying the user data flow are cleared, but the PDP context still remains active and the signalling connection is also maintained.

In the case of UE-terminated packet data the GGSN initiates the procedure by sending a data packet to the SGSN currently serving the addressed UE. The receipt of this data packet causes the SGSN to send packet paging to the desired UE. Packet paging forces the UE to change the RRC state from "connected URA PCH" to "connected cell FACH" and the UE sends a "service request" message to the network. Since SM is in active state the network is able to allocate RAB and CN bearers according to the correct, negotiated parameters describing the packet connection. When the bearers have been allocated, the packet data are transferred to the UE, and if the UE has packet data to be sent they are sent to the network. When the packet data have been transferred the RAB and CN bearers carrying the user data flow are cleared and only signalling bearers remain.

When the subscriber switches his or her terminal off, "PDP context deactivation" takes place. This procedure removes any address information stored in the network concerning the packet connection and the PDP context is deleted. This causes SM to change its state from active to inactive and packet data transfer is no longer possible. Since the UE switches itself off, the signalling connection is no longer required and is thus released. As a consequence, the PMM state goes to "detached" and the RRC state changes to "idle".

6.3 Charging, Billing and Accounting

This section provides a short overview of the charging, billing and accounting mechanisms used and their possible use in UMTS networks. But, first, these terms and their meanings should be clarified:

- *Charging* is a collection of procedures generating charging data. These procedures are located in CN elements. Charging (i.e., the identification of collected data) is specified on a common level in UMTS specifications.
- *Billing* is a procedure that post-processes charging data and, as a result, produces a bill for the end-user. Billing as such is beyond the scope of the UMTS specification. Instead of specifications, local laws and marketing practices regulate billing.
- *Accounting* is a common name for charging data collected over a predefined time period. The difference between billing and charging is that in the former accounting information is collected from the connections *between* operators or various commercial bodies. Hence, accounting has nothing directly to do with end-users. Another point is that, as far as telecommunication is concerned, accounting and charging are different issues, but in the context of the Internet these two terms are often used synonymously.

6.3.1 Charging and Accounting

As a result of its historical background and the inherited nature of UMTS networks, the network must support three triggers for end-user traffic identification and charging purposes. These three triggers are:

- *Time-based*: the system collects information about transaction duration: when it started, when it finished and how long it took.
- *Quantity-based*: the system collects information about the number of bits transferred during the transaction.
- *Quality-based*: the system collects information about the quality criteria used during the transaction. The quality profile/criteria of the transaction is called the "Quality of Service" (QoS). The QoS, its parameters and mechanisms are briefly described in Chapter 8.

Of these, the first was traditionally used in CS environments (e.g., PSTNs). The other two were introduced in cellular networks a couple of years ago in the context of the GPRS. Since the UMTS offers a number of transaction possibilities, these triggers may not be adequate. Note also that as the networking model has evolved it has become more complex and there are a number of commercial bodies involved. It is for these reasons that the charging requirements can be shortlisted (list is not exhaustive, though):

- Charging must be able to be applied separately for each medium type (voice, video, data) within a session and also for each used service (call, streaming video, file download, etc.).
- Charging must be able to be applied separately for the various QoS levels allocated for the medium or services within a session.
- It must be possible to charge each leg of a session or a call separately. This includes incoming and outgoing legs and any forwarded/redirected legs. The legs mentioned here are logical legs (i.e., not necessarily identical to the actual signal and traffic flow).
- Charging can be based on the access method used (i.e., 2G, 3G or complementary access). On the other hand, the operator may select access-independent charging and charge for actual service usage.
- It must be possible for the home network to charge its customers while roaming in the same way as when they are at home. For example, if duration-based charging is used for streaming music in the home network, then it must be possible to apply the same principle when the user is roaming.
- It must be possible for operators to have the option of applying charging mechanisms that are used in the GSM/GPRS, such as duration of a voice call, the amount of data transmitted (e.g., for streaming, file download, browsing) and for an event (one-off charge).
- It must be possible for charging to be applied based on location, presence, push services, etc.
- It must be possible to charge using pre-pay, post-pay, advice of charge and third-party charging techniques.
- It must be possible for the home network to apply different tariffs to national calls and short messages established/sent by their subscribers while roaming in their home PLMN, irrespective of whether or not the called subscriber's Home PLMN equals the calling subscriber's home PLMN, rather than on the called subscriber's MSISDN.

When these requirements are combined in a commercial environment, the result will be like that shown in Figure 6.24.

As can be seen, there are many commercial bodies involved. So far in this section, we have concentrated on "retail charging" and identified its requirements. Retail charging is identical to the term "charging" we defined at the beginning of this section. The other charging types identified in Figure 6.24 are basically accounting (i.e., they are not directly visible to end-users).

A typical accounting case is "wholesale charging" where a virtual operator/service provider buys network capacity and sells this capacity to its subscribers. This business is becoming more common nowadays. In many countries the authorities regulate prices used in accounting interfaces to guarantee competition. A good example of this is the case in Finland where local authorities ensure that mobile network operators have

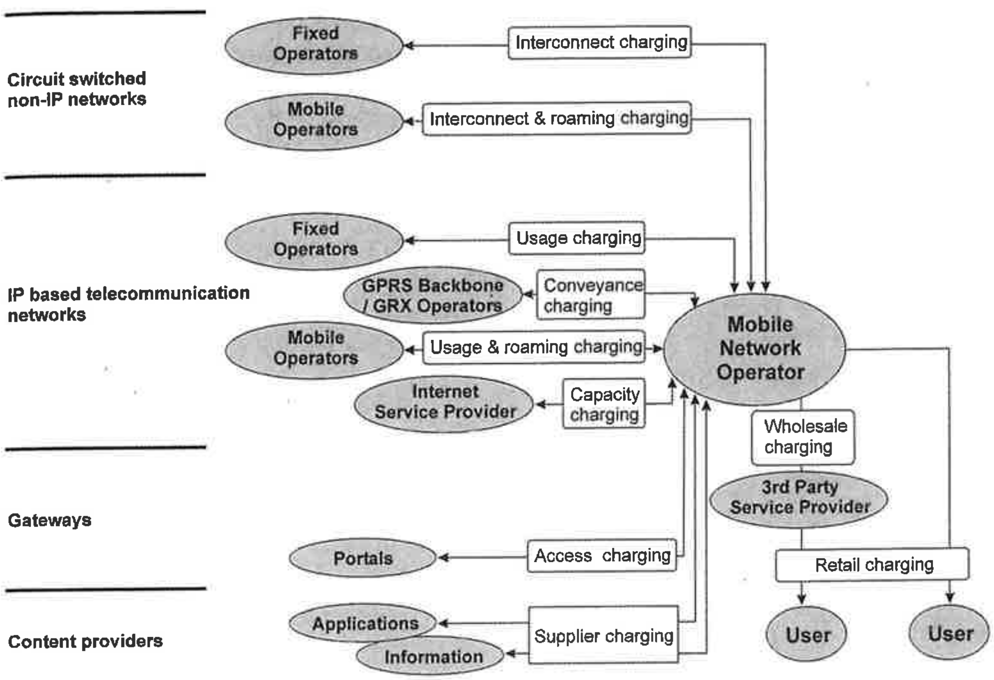


Figure 6.24 Charging types as defined in 3GPP TS 22.115

common and equal pricing for so-called virtual operators (i.e., any virtual operator can in principle buy network capacity from any mobile network operator).

In addition to "wholesale charging" the mobile network operator has altogether four further accounting interfaces: non-IP based telecommunication networks, IP-based telecommunication networks, various gateway types of services and content providers. Non-IP based telecommunication networks are mostly CS in nature and the interconnection towards these networks can be done in two ways. If the CN contains a CS domain, it interconnects directly using methods already defined in 3GPP R99. This means the control plane is taken care of by SS7 and ISUP types of signalling and the user plane is a timeslot(s) on a PCM trunk. If the CS domain follows 3GPP R4 implementation, the user plane is interconnected through CS-MGW elements and any related signalling is handled by means of signalling gateway functionality. In both cases the collected accounting information shows CS resource usage and charging is mainly based on call duration. For IP-based telecommunication networks the accounting information contains session information: what kinds of sessions and what transactions were made during the session? As far as gateways (portals) are concerned, the collected accounting information contains accessing figures (i.e., how often a portal is used). While, for content providers, the accounting information describes how many times a certain content is accessed, when it was accessed and by whom.

CS transactions and their charging will remain as it is and no major changes are expected in this area. PS communication in turn will change and its charging will be very challenging. PS communication is handled using the IMS, whose architectural aspects are handled in Section 6.4. From the charging and accounting point of view the IMS will involve the majority of these briefly presented accounting interfaces as well as retail charging.

Table 6.1 presents the various available charging options. In the table, A, B and C are parties to multimedia transactions performed through the IMS. As stated, the table presents charging *options*, rather than real charging cases. Finally, collected charging and accounting data is up to the IMS operator as is final billing, but the options presented in the table should be supported in the IMS configuration.

A multimedia session consists of media *components*. According to the requirements illustrated at the beginning of this section, the network must be able to recognise various media components and handle charging and accounting data accordingly.

The recognised media components are:

- Voice.
- Real-time audio.
- Streaming audio.
- Real-time video.
- Streaming video.
- Data download/upload.
- Interactive data (e.g., Web-browsing).
- Messaging (SMS, MMS type).
- Email.
- Data stream with unspecified content; this is where the network operator acts as a "bit pipe" and charging is based on quantity (bits transferred) and quality (QoS profile used in the transaction).

Table 6.2 summarises the available charging mechanism and type options for different media components.

6.3.2 Billing

Billing is not part of the scope of UMTS specifications: it is a separate process implemented using separate equipment foreign to the UMTS network. The basis of billing is the charging and accounting data collected from the operator's own network and, possibly, received from other networks.

Billing as a process is regulated by local authorities and laws and, thus, billing principles vary from country to country. In addition to charging and accounting information the billing process requires other source data:

- *Subscription information*. This defines the business relationship between the operator/service provider and end-user: provisioned services, QoS profiles, identities, etc.
- *Interconnection/Service Level Agreement (SLA)*. This defines the business relationship between operators, service providers and content providers: bandwidth, QoS guarantees, O&M-related issues, etc.
- *Pricing policy*. This is defined by the operator/service provider and is regulated by

Table 6.1 Charging options available in multimedia sessions

No.	Connection	Description	Charging options required
1	*A* sets up a session with *B*	A simple connection between two subscribers or a subscriber and a service (e.g., voicemail)	*A* pays for the session set-up with *B* *A* pays for the session resource with *B* *B* pays for the session resource with *A*
2	*A* sets up a session with *B*	A simple connection where *B* is a "toll-free" (800)-type service	*B* pays for the session set-up *B* pays for the session resource *A* pays for part of the session resource (i.e., allowing split charging between *A* & *B*)
3	*A* requests a session with *B*, *B* redirects to *C*	This is redirection. The connection path is not set up to *B* from *A*, instead *A* is told to set up a connection direct to *C*	*A* pays for the session set-up with *B* *A* pays for the session resource with *C* *C* pays for the session resource with *A* *A* pays for the session resource as though it were with *B* and *B* pays for the session resource with *C* as though it came from *B*
4	*A* requests a session with *B*, *B* forwards it to *C*	This is normal forwarding as in the GSM. The connection path is *A* to *B*'s home network and *B*'s home network to *C*	*A* pays for the session set-up with *B* *A* pays for the session resource as though it were with *B* and *B* pays for the session resource with *C*
5	*A* sets up sessions with multiple parties (multi-party)	Connections to multiple parties are initiated by *A*	*A* pays for the set-up of each session *A* pays for each of the session resources with each of the called parties Each of the called parties pays for the session resource with *A*

(continued)

Table 6.1 (*cont.*)

6	*A* has a multi-party session where the individual parties set up the session with A	The multiple parties in the session initiate the session with *A*	Each party pays for the session set-up with *A* *A* pays for the session resource with the multiple parties The individual parties in the session each pay for the session resource with *A*
7	*A* is in a session with *B*, then puts *B* on hold to set up a session with *C*, then returns to *B* after dropping *C*	*A* still has a connection with *B* while also in a session with *C*. The session with *B* continues after the session with *C* is terminated	*A* pays for each session set-up with *B* & *C* *A* pays for the session resource with *B* & *C* *B* & *C* pay for the session resource with *A*
8	*A* is in a session with *B* then answers a session request from *C* while keeping *B* on hold	*A* still has a connection with *B* while also in a session with *C*. The session with *B* continues after the session with *C* is terminated	*A* pays for the session set-up with *B* *C* pays for the session set-up with *A* *A* pays for the session resource with *B* & *C* *B* & *C* pay for the session resource with *A*
9	*A* sets up a session with *B* who is roaming in another network	The connection is made from *A* to *B*'s home network and then forwarded to *B* in the visited network (normal GSM mechanism). Alternatively, *A* is redirected directly to *B* in the	*A* pays for the session set-up with *B* *A* pays for the session resource as though it were with *B* in his home network and *B* pays for the session resource from his home network with the visited network *A* pays for the session resource with *B* in the visited network *B* pays for the session resource with *A*

Table 6.2 Charging mechanism and type options

Component	Charging mechanism options	Charging-type options
Voice	Charging principles as described in Table 6.1	Charging by duration of session Charging by QoS requested and/or delivered One-off set-up charge
Real time audio and video	Charging principles as described in Table 6.1	Charging by duration of session Charging by QoS requested and/or delivered One-off set-up charge
Streaming audio and video	Charged to the initiator of the request Charged to the sender of the audio or video	Charging by duration of session Charging by volume of data, optionally QoS-differentiated One-off set-up charge
Data (upload or download)	Charged to the initiator of the request Charged to the sender of the data	Charging by duration of session Charging by volume of data, optionally QoS-differentiated One-off set-up charge
Interactive data	Charged to the initiator of the session	Charging by duration of session Charging by volume of data, optionally QoS-differentiated One-off set-up charge
Messaging (SMS, MMS type)	Charged to the initiator of the message Charged to the recipient of the message	Charging by event (e.g., SMS) Charging by volume of data
Unspecified content (data stream)	Charged to the initiator of the session Charged to all parties involved	Charging by duration of session Charging by volume of data (sent & received), optionally QoS-differentiated One-off set-up charge

the marketplace and in some cases by authorities: price per call, price per session, price per media component, monthly fees, reduced tariffs, etc.

The function of billing is to combine these items with charging and accounting data and to produce bills for end-users and other related parties.

6.4 IP Multimedia Subsystem (IMS)

The previous sections in this chapter covered CN domain structures and CN management tasks and controlling duties. This section will cover the final ingredient of the whole architecture, the IP Multimedia Subsystem (IMS), which enables applications in mobile devices to establish peer-to-peer connections.

People have a natural need to share experiences: share what they see, share things they do, share emotions. Nowadays people have plain telephony to talk to each other, the Multimedia Messaging Service (MMS) to send pictures and voice clips and the possibility of browsing web pages using their terminals. Some may think that this is sufficient, but there is a wave of other multimedia communication services coming, such as interactive gaming, interactive web services, application sharing, video communication, rich messaging, presence and group communication (e.g., Push to talk over Cellular, or PoC). Clearly, many of these new services will be exercised simultaneously.

UMTS networks bring flexible IP bearers and excellent data capabilities to terminals using the GPRS, Enhanced Data for GSM Evolution (EDGE) and Wideband Code Division Multiple Access (WCDMA) networks. However, the network lacks a mechanism to connect terminals using IP. This is where the IMS comes in. As shown in Figure 6.25 the IMS introduces multimedia session control using SIP in the PS domain; this allows users to establish connections with various ASs and, especially, to use the IP-based services between the terminals.

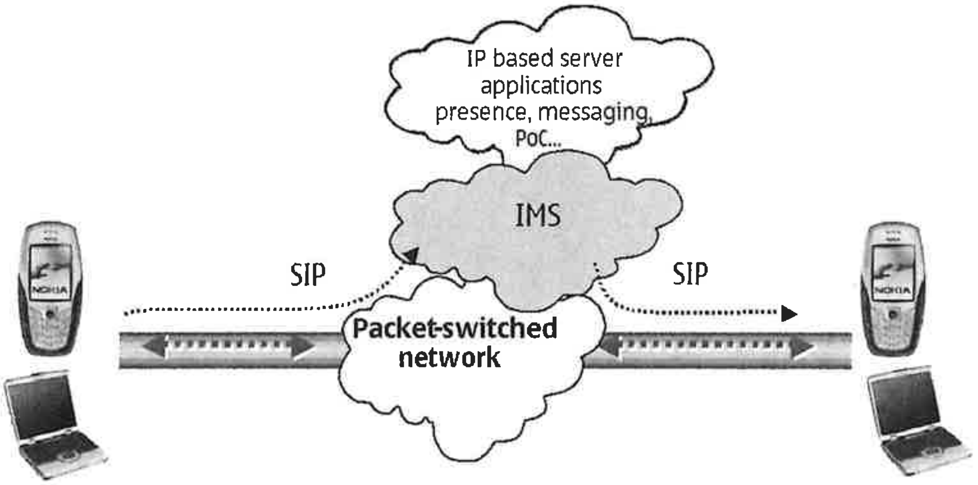


Figure 6.25 IMS brings multimedia session control in the Packet Switched (PS) domain

In this section we introduce the IMS. We will explain IMS design principles and you will also get a grip on the building blocks of the IMS, how different functions are connected and the key protocols of the IMS. However, this chapter is intentionally light on examples and protocol behaviour. You will find a detailed and complete description of the IMS in the book *IMS IP Multimedia Concepts and Services in the Mobile Domain.*

6.5 IP Multimedia Subsystem Fundamentals

There is a set of basic requirements that guides the way in which the IMS architecture has been created and how it should evolve in the future. The following ten issues form the baseline for the IMS architecture:

- IP connectivity.
- Access independence.
- Layered design.
- Quality of Service (QoS).
- IP policy control.
- Secure communication.
- Charging.
- Possibility to roam.
- Interworking with other networks.
- Service development and service control for IP-based applications.

As the name "IP Multimedia Subsystem" implies, a fundamental requirement is that a terminal has to have IP connectivity to access it. Peer-to-peer applications require end-to-end reachability and this connectivity is easiest attainable with IP version 6 (IPv6) because IPv6 does not have address shortage. Therefore, the 3GPP has arranged matters so that the IMS exclusively supports IPv6 [3GPP TS 23.221]. However, early IMS implementations and deployments may use IP version 4 (IPv4). There exists a study report that contains guidelines and recommendations if IPv4 is used to access the IMS [3GPP TR 23.981]. IP connectivity can be obtained either from the home network or the visited network when a user is roaming. The leftmost part of Figure 6.26 presents an option in which the UE has obtained an IP address from a visited network. In the UMTS network this means that the Radio Access Network (RAN), SGSN and GGSN are located in the visited network. The rightmost part of Figure 6.26 presents an option in which the UE has obtained an IP address from a home network. In the UMTS network this means that the RAN and SGSN are located in the visited network. Obviously, when a user is located in the home network all necessary functions are in the home network and IP connectivity is obtained in that network.

Although this is a book about UMTS networks it is important to realise that the IMS is designed to be access-independent so that IMS services can be provided over any IP connectivity networks (e.g., GPRS, WLAN, broadband access x-Digital Subscriber Line). In fact the first IMS release, Release 5, is tied to UMTS because the only possible

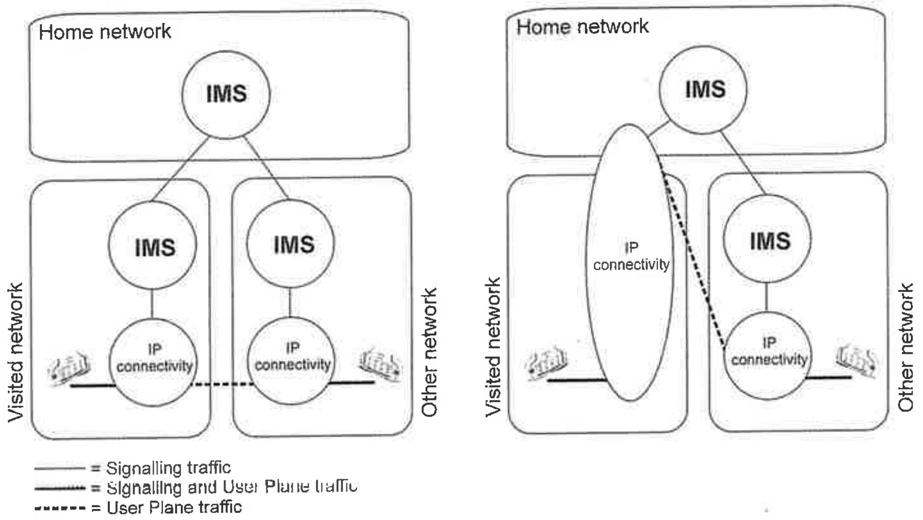
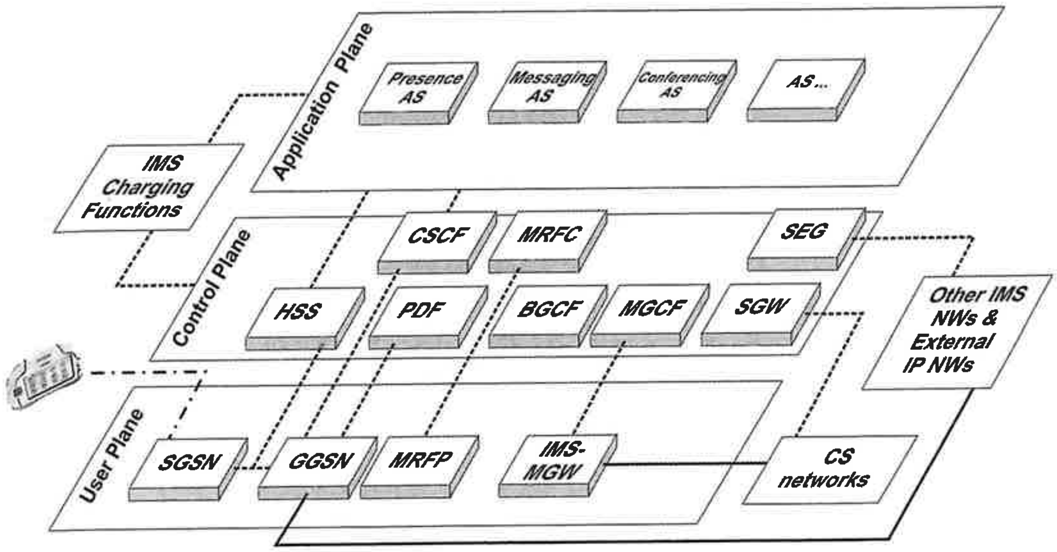


Figure 6.26 IP Multimedia Subsystem (IMS) connectivity options when a user is roaming

IP connectivity access network is GPRS. However, the specifications were corrected in the second release, Release 6, in such a way that all other accesses are possible as well.

In addition to access independence the IMS architecture is based on a layered approach. This means that transport and bearer services are separated from the IMS signalling network and session management services. Further services are run on top of the IMS signalling network (Figure 6.27 shows the design). The layered approach aims at a minimum dependence between layers. A benefit is that it facilitates the addition of new access networks to the system later on. Wireless Local Area Network (WLAN) access to the IMS, in 3GPP Release 6, will test how well the layering has been done. Other accesses may follow (e.g., fixed broadband). The layered approach increases the importance of the application layer. When applications are isolated and common functionalities can be provided by the underlying IMS network the same applications can run on the UE using diverse access types.

Quality of Service (QoS) plays an important role in IP networks, as it is commonly known that delays tend to be high and variable, packets arrive out of order and some packets are completely lost on the public Internet. This paradigm must be corrected in the IMS otherwise end-users will not make good use of the new, fancy IMS services. Via the IMS, the UE negotiates its capabilities and expresses its QoS requirements during a SIP session set-up or session modification procedure. The UE is able to negotiate such parameters as: media type, direction of traffic, media bit rate, packet size, packet transport frequency. After negotiating the parameters at the application level, the UE is able to map information to UMTS QoS parameters and is able to reserve suitable resources from the RAN and GPRS network. Chapter 8 in this book contains information about QoS classification, attributes and related network mechanisms. We have just described how UEs could guarantee QoS between the UE and the GGSN for completing end-to-end QoS. In addition, we need the connections between the GGSNs to provide the necessary QoS. This could be achieved by SLAs between operators.



┈┈┈┈┈ = Signalling or controlling traffic
——— = User Plane traffic
—·—·—· = Both

Note: connections inside the plane and User Plane connections to/from AS are not shown.

Figure 6.27 IP Multimedia Subsystem (IMS) and layering architecture

IP policy control means the capability to authorise and control the usage of bearer traffic intended for IMS media, based on the signalling parameters at the IMS session. This requires interaction between the GPRS network and the IMS:

- The policy control element is able to verify that values negotiated in SIP signalling are used when activating bearers for media traffic. This allows an operator to verify that its bearer resources are used for that particular peer-to-peer connection as was negotiated in SIP signalling.
- The policy control element controls when media traffic between the end points of a SIP session can start or stop.
- The policy control element is able to receive notifications from the GPRS network about modification, suspension or deactivation of the PDP context(s) of a user associated with a SIP session.

Security is a fundamental requirement in every part of the UTMS network and the IMS is not an exception. The IMS has its own authentication and authorisation mechanisms between the UE and the IMS network in addition to the GPRS network procedures. Moreover, the integrity and *optional* confidentiality of SIP messages is provided between the UE and the IMS network and between IMS network elements regardless of the underlying RAN and GPRS network. The security issues in the UMTS environment are described in detail in Chapter 9.

From an operator or service provider perspective, the ability to charge users is a must in any network. The IMS architecture allows different charging models to be used. This includes, say, capability to charge just the calling party or charge both the calling party and the called party based on resources used at the transport level. In the latter case the calling party could be charged entirely for an IMS-level session: that is, it is possible to use different charging schemes at the transport and IMS level. However, an operator might be interested in correlating charging information generated at the transport and IMS (service and content) charging levels. This capability is provided if an operator utilises a policy control reference point. As IMS sessions may include multiple media components (e.g., audio and video), it is required that the IMS provide a means for charging per media component. This would allow the possibility of charging the called party if he or she adds a new media component in a session. It is also required that different IMS networks are able to exchange information about the charging to be applied to a current session.

From a user point of view, it is important to gain access to his or her services regardless of geographical location. Roaming is a feature that allows the use of services when the user is not geographically located in the service area of the home network. Three different types of roaming models can be identified: GPRS roaming, IMS roaming and IMS CS roaming. GPRS roaming entails the capability of accessing the IMS when a visited network provides the RAN and SGSN, and the home network provides the GGSN and IMS. The IMS roaming model refers to a network configuration in which the visited network provides the RAN, SGSN, GGSN and the IMS entry point (i.e., P-CSCF) and the home network provides the remaining IMS functionalities. Roaming between the IMS and the CS CN domain refers to inter-domain roaming between the IMS and CS CN. This means that if a user is not reachable in the IMS, then the IMS routes the session towards the CS CN and vice versa.

It is foreseen that different types of networks will co-exist for many years; therefore, inter-working with legacy networks (PSTN, ISDN, mobile, Internet) is an important aspect of any new network architecture. Moreover, it is probable that some people will not be willing to switch terminals or subscriptions to adopt new technology innovations, but there again some users may have more than one item of UE with totally different capabilities: WLAN-enabled UE for office usage, UMTS UE for outdoor usage, wire line UE for home. This will raise the issue of being able to reach people regardless of what kind of terminals they have or where they are located. To be a new, successful communication network technology and architecture the IMS has to be able to connect as many users as possible. Therefore, the IMS supports communication with PSTN, ISDN, mobile and Internet users.

The CS CN and the IMS differ regarding the service control model and service deployment. In the CS CN the visited service control is in use. This means that, when a user is roaming, an entity in a visited network provides services and controls the traffic for the user. This entity is called a Visited Mobile Switching Centre (VMSC). In contrast, home service control is in use in the IMS, meaning that Serving-Call Session Control Function (S-CSCF), which is always located in the home network, controls services. The importance of having a scalable service platform and the possibility to launch new services rapidly has meant that the old way of standardising complete sets of teleservices, applications and supplementary services no longer exists. This is the reasons the IMS provides a service framework that provides the necessary capabilities to support various multimedia applications within the IMS and on top of the IMS. Actually, the IMS is not a service itself; rather, it is a SIP-based architecture for enabling an advanced IP service and application on top of the PS network. The IMS provides the necessary means for invoking services. Examples of these kinds of applications are presence and conferencing. In the future, it is expected that the Open Mobile Alliance (OMA) will make maximum use of the IMS while developing various applications and services.

6.6 IMS Entities and Functionalities

This section discusses IMS entities and key functionalities. These entities can be roughly classified in six main categories:

- Session management and routing family (CSCFs).
- Databases (HSS, SLF).
- Inter-working functions (BGCF, MGCF, IMS-MGW, SGW).
- Services (AS, MRFC, MRFP).
- Support functions (THIG, SEG, PDF).
- Charging.

It is important to understand that IMS standards are set up such that the internal functionality of network entities is not specified in detail. Instead, standards describe reference points between entities and functionalities supported at the reference points. A good example is: How does the CSCF obtain user data from databases?

6.6.1 Call Session Control Functions (CSCFs)

There are three different kinds of Call Session Control Functions (CSCFs): Proxy-CSCF (P-CSCF), Serving-CSCF (S-CSCF) and Interrogating-CSCF (I-CSCF). Each CSCF has its own special tasks (these tasks are described in the following sections). All CSCFs play a role during registration and session establishment and form the SIP routing machinery. Moreover, all functions are able to send charging data to an offline charging function. There are some functions common to the P-CSCF and S-CSCF that they can perform. Both entities are able to release sessions on behalf of the user (e.g., when the S-CSCF detects a hanging session or the P-CSCF receives a notification that a media bearer is lost) and are able to check the content of Session Description Protocol (SDP) payload and to check whether it contains media types or codecs that are not allowed for a user. When the proposed SDP does not fit the operator's policy, the CSCF rejects the request and sends a SIP error message to the UE.

6.6.1.1 Proxy-Call Session Control Function (P-CSCF)

The P-CSCF is the first contact point for users within the IMS. This means that all SIP signalling traffic from the UE will be sent to the P-CSCF. Similarly, all terminating SIP signalling from the network is sent from the P-CSCF to the UE. There are four unique tasks assigned to the P-CSCF: SIP compression, IP security (IPSec) association, interaction with the Policy Decision Function (PDF) and emergency session detection.

As SIP is a text-based signalling protocol, it contains a large number of headers and header parameters, including extensions and security-related information, which means that SIP message sizes are larger than those of binary encoded protocols. In the IMS architecture, setting up a SIP session is a tedious process involving codec and extension negotiations as well as QoS inter-working notifications. In general, this provides a flexible framework that allows sessions with differing requirements to be set up. However, the drawback is the large number of bytes and messages exchanged over the radio interface. The increased message size means that:

- Session set-up procedures when using SIP will take much more time to be completed than those using existing cell-specific signalling, which means that the end-user will experience a delay in session establishment that will be unexpected and likely unacceptable.
- Intra-call signalling will in some way adversely affect voice quality/system performance.

To speed up session establishment the 3GPP has mandated the support of SIP compression. The UE has a means of indicating to the P-CSCF that it wants to receive signalling messages compressed over the air interface.

The P-CSCF is responsible for maintaining Security Association (SA) and applying integrity and confidential protection of SIP signalling. This is achieved during SIP registration when the UE and P-CSCF negotiate IPSec SAs. After initial registration the P-CSCF is able to apply integrity and confidentiality protection to SIP signalling (Chapter 9 describes IPSec in more detail).

The P-CSCF is tasked with relaying session- and media-related information to the PDF when an operator wants to apply IP policy control. Based on the received information the PDF is able to derive authorised IP QoS information that will be passed to the GGSN when the GGSN needs to perform IP policy control prior to accepting a secondary PDP context activation. Moreover, via the PDF the IMS is able to deliver IMS-charging correlation information to the GPRS network, and, similarly, via the PDF the IMS is able to receive GPRS-charging correlation information from the GPRS network. This makes it possible to merge charging data records coming from the IMS and GPRS networks in the billing system.

IMS emergency sessions are not yet fully specified; therefore, it is a prerequisite that the IMS network detect emergency session attempts and guide a UMTS UE to use the CS network for emergency sessions. Detection is the task of the P-CSCF. This functionality will not vanish when IMS emergency sessions are supported because in certain roaming cases it is possible that the UE does not itself recognise that it has dialled an emergency number.

6.6.1.2 Interrogating-Call Session Control Function (I-CSCF)

The I-CSCF is a contact point within an operator's network for all connections destined for a subscriber of that network operator. There are four unique tasks assigned to the I-CSCF:

- Obtaining the name of the S-CSCF from the HSS.
- Assigning an S-CSCF based on received capabilities from the HSS. The assignment of the S-CSCF will take place when a user registers with the network or a user receives a SIP request while unregistered but still wants to receive services related to the unregistered state (e.g., voicemail).
- Routing incoming requests further to an assigned S-CSCF.
- Providing Topology Hiding Inter-network Gateway (THIG) functionality (THIG is further explained in Section 6.6.5).

6.6.1.3 Serving Call Session Control Function

The S-CSCF is the focal point of the IMS because it is responsible for handling the registration process, making routing decisions, maintaining session states and storing the service profile.

When a user sends a registration request it will be routed to the S-CSCF, which downloads authentication data from the HSS. Based on the authentication data, it then generates a challenge to the UE. After receiving the response and verifying it the S-CSCF accepts the registration and starts supervising the registration status. After this procedure the user is able to initiate and receive IMS services. Moreover, the S-CSCF downloads a service profile from the HSS as part of the registration process.

A service profile is a collection of user-specific information that is permanently stored in the HSS. The S-CSCF downloads the service profile associated with a particular public user identity (e.g., *joe.doe@ims.example.com*) when this particular public user identity is registered with the IMS. The S-CSCF uses information included in the service profile, which decides when and, in particular, which AS(s) is contacted when a user sends a SIP request or receives a request from somebody. Moreover, the service profile may contain further instructions about the kind of media policy the S-CSCF need to apply; for example, it may indicate that a user is only allowed to use audio and application media components but not video media components.

The S-CSCF is responsible for key routing decision as it receives all UE-originated and UE-terminated sessions and transactions. When the S-CSCF receives a UE-originating request via the P-CSCF it needs to decide whether ASs need to be contacted prior to sending the request further on. After possible AS(s) interaction the S-CSCF either continues a session in the IMS or breaks to other domains (CS or other IP networks). Moreover, if the UE uses an MSISDN number to address a called party, then the S-CSCF converts the MSISDN number (i.e., a tel URL) to SIP URI format prior to sending the request further because the IMS does not route requests based on MSISDN numbers. Similarly, the S-CSCF receives all requests that will terminate at the UE. Although the S-CSCF knows the IP address of the UE from the registration it routes all requests via the P-CSCF because the P-CSCF takes care of SIP compression

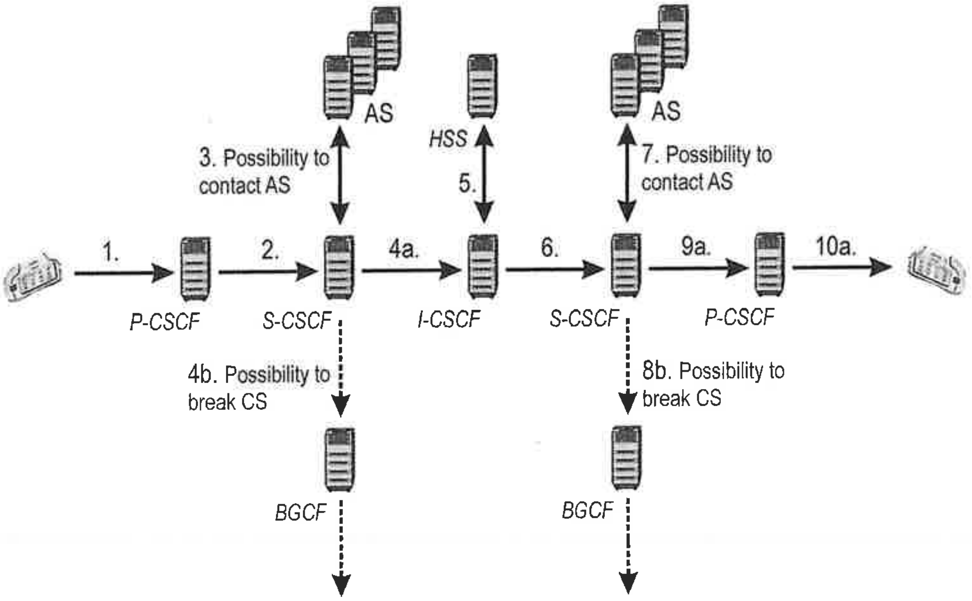

AS
HSS
AS
3. Possibility to contact AS
5.
7. Possibility to contact AS
1.
2.
4a.
6.
9a.
10a.
P-CSCF
S-CSCF
I-CSCF
S-CSCF
P-CSCF
4b. Possibility to break CS
8b. Possibility to break CS
BGCF
BGCF

Figure 6.28 Serving-Call Session Control Function (S-CSCF) routing and basic IP Multimedia Subsystem (IMS) session set-up

and security functions. Prior to sending a request to the P-CSCF, the S-CSCF may route the request to an AS(s) (e.g., to check possible redirection instructions). Figure 6.28 illustrates the S-CSCF's role in routing decisions.

6.6.2 Databases

There are two main databases in the IMS architecture: Home Subscriber Server (HSS) and Subscription Locator Function (SLF).

The HSS is the main data storage for all subscriber- and service-related data of the IMS. The main data stored in the HSS include user identities, registration information, access parameters and service-triggering information [3GPP TS 23.002]. User identities consist of two types: private and public user identities (see Sections 6.2.1.1.7 and 6.2.1.1.8). Private user identity is a user identity that is assigned by the home network operator and is used for such purposes as registration and authorisation, while public user identity is an identity that other users can use for requesting communication with the end-user. IMS access parameters are used to set up sessions and include parameters like user authentication, roaming authorisation and allocated S-CSCF names. Service-triggering information enables SIP service execution. The HSS also provides user-specific requirements for S-CSCF capabilities. This information is used by the I-CSCF to select the most suitable S-CSCF for a user. In addition to functions related to IMS functionality, the HSS contains the subset of the HLR/AuC functionality required by the PS domain and the CS domain. Communication between different HSS functions is not standardised. There may be more than one HSS in a home network depending on the number of mobile subscribers, the capacity of the equipment and the organisation of the network.

The SLF is used as a resolution mechanism that enables the I-CSCF, the S-CSCF and the AS to find the address of the HSS that holds the subscriber data for a given user identity when multiple and separately addressable HSSs have been deployed by the network operator.

6.6.3 *Interworking Functions*

This section introduces the four interworking functions that are needed for exchanging signalling and media between the IMS and CS CN.

Section 6.6.1.3 explained that it is the S-CSCF that decides when to break to the CS CN. To do this the S-CSCF sends a SIP session request to the Breakout Gateway Control Function (BGCF), which chooses where a breakout to the CS domain should occur. The outcome of a selection process can be either a breakout in the same network in which the BGCF is located or another network. If the breakout happens in the same network, then the BGCF selects a Media Gateway Control Function (MGCF) to handle the session further. If the breakout takes place in another network, then the BGCF forwards the session to another BGCF in the selected network [3GPP TS 23.228]. The latter option allows routing of signalling and media over IP near to the called user.

When a SIP session request hits the MGCF it performs protocol conversion between SIP protocols and ISUP (or the Bearer Independent Call Control, BICC) and sends a converted request via the Signalling Gateway (SGW) to the CS CN. The SGW performs signalling conversion (both ways) at the transport level between the IP-based transport of signalling (i.e., between Sigtran SCTP/IP and SS7 MTP) and the Signalling System #7 (SS7)-based transport of signalling. The SGW does not interpret application layer (e.g., BICC, ISUP) messages. The MGCF also controls the IMS Media Gateway (IMS-MGW). The IMS-MGW provides the user plane link between CS CNs and the IMS. It terminates the bearer channels from the CS CN and media streams from the backbone network (e.g., RTP streams in an IP network or AAL2/ATM connections in an ATM backbone), executes the conversion between these terminations and performs transcoding and signal processing for the user plane when needed. In addition, the IMS-MGW is able to provide CS users with tones and announcements.

Similarly, all incoming call control signalling from IMS users is destined to the MGCF that performs the necessary protocol conversion and sends a SIP session request to the I-CSCF for session termination. At the same time, the MGCF interacts with the IMS-MGW and reserves the necessary IMS-MGW resources at the user plane.

Figure 6.29 visualises the inter-working concept. The leftmost part of the figure presents an IMS-originated session and the rightmost part of the figure shows a CS-originated call.

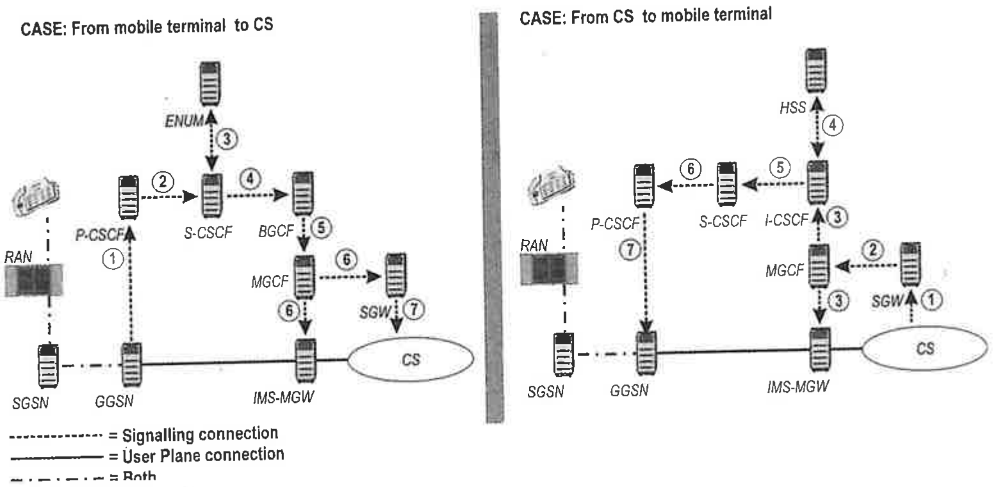


Figure 6.29 IP Multimedia Subsystem (IMS) and Circuit Switched (CS) inter-working

6.6.4 Service-related Functions

Three functions in this book are categorised as IMS service-related functions: Multimedia Resource Function Controller (MRFC), Multimedia Resource Function Processor (MRFP) and AS.

Keeping the layered design in mind, ASs are not pure IMS entities; rather, they are functions on top of IMS. However, ASs are described here as part of IMS functions because ASs are the entities that provide value-added multimedia services in the IMS, such as presence and PoC. An AS resides in the user's home network or in a third-party location. "Third party" here means a network or a stand-alone AS. The main functions of the AS are:

- The possibility to process and have an impact on an incoming SIP session received from the IMS.
- The capability to originate SIP requests.
- The capability to send accounting information to the charging functions.

The services offered are not limited purely to SIP-based services, since an operator is able to offer access to services based on the CAMEL Service Environment (CSE) and the Open Service Architecture (OSA) for its IMS subscribers [3GPP TS 23.228]. Therefore, "AS" is the term generically used to capture the behaviour of the SIP AS, OSA Service Capability Server (SCS) and CAMEL IMS Switching Function (IMS-SF).

Figure 6.30 shows how the different functions are connected. From the perspective of the S-CSCF SIP AS, the OSA service capability server and the IMS-SF exhibit the same reference point behaviour. An AS may be dedicated to a single service and a user may have more than one service; therefore, there may be one or more ASs per user. Additionally, there may be one or more ASs involved in a single session. For example, an operator could have one AS to control the terminating traffic to a user

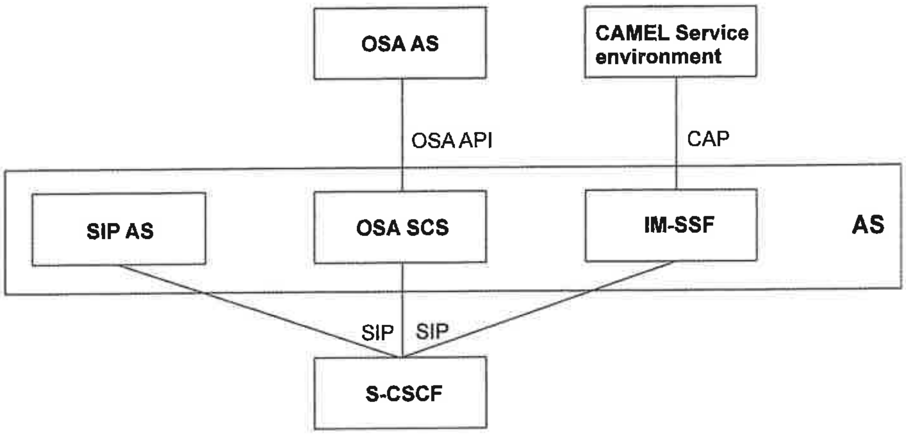


Figure 6.30 Relationship between different Application Server (AS) types

based on user preferences (e.g., redirecting all incoming multimedia sessions to an answer machine between 5 p.m. and 7 a.m.) and another AS to adapt the content of instant messages according to the capabilities of the UE (screen size, number of colours, etc.).

The MRFC and MRFP together provide mechanisms for bearer-related services, such as conferencing, announcements to a user or bearer transcoding in the IMS architecture. The MRFC is tasked to handle SIP communication to and from the S-CSCF and to control the MRFP. The MRFP in turn provides the user plane resources that are requested and instructed by the MRFC. The MRFP performs the following functions:

- Mixing of incoming media streams (e.g., for multiple parties).
- Media stream source (for multimedia announcements).
- Media stream processing (e.g., audio transcoding, media analysis) [3GPP TS23.228, TS 23.002].

Currently, the role of MRFC and MRFP in the IMS architecture is minor because the MRFC is co-located with an AS in the IMS conferencing work [3GPP TS 24.147] and the reference point between the MRFC and MRFP is not yet well defined.

6.6.5 Support Functions

Separation of the control plane and the user plane was maybe one of the most important issues of IMS design. Full independence of the layers is not feasible because, without interaction between the user plane and the control plane, operators are not able to control QoS, the source/destination of IMS media traffic and when the media starts and stops. Therefore, a mechanism to authorise and control the usage of the bearer traffic intended for IMS media traffic was created; this is based on the SDP parameters negotiated at the IMS session. This overall interaction between the GPRS and the

IMS is called a "Service Based Local Policy" (SBLP). It was later that the additional capability of exchanging charging correlation information was specified. The PDF is responsible for making policy decisions based on session- and media-related information obtained from the P-CSCF. It acts as a policy decision point for SBLP control.

Session establishment in the IMS involves end-to-end message exchange using SIP and SDP. During message exchange the UEs negotiate a set of media characteristics (e.g., common codec(s)). If an operator applies the SBLP, then the P-CSCF will forward the relevant SDP information to the PDF together with an indication of the originator. Correspondingly, the PDF allocates and returns an authorisation token which the P-CSCF will pass on to the UE. The PDF notes and authorises the IP flows of the chosen media components by mapping from SDP parameters to authorised IP QoS parameters for transfer to the GGSN via the Go interface. When the UE activates or modifies a PDP context for media it has to perform its own mapping from SDP parameters and application demands to some UMTS QoS parameters. PDP context activation or modification will also contain the received authorisation token and flow identifiers as the binding information. On receiving the PDP context activation or modification, the GGSN asks for authorisation information from the PDF. The PDF compares the received binding information with the stored authorisation information and returns an authorisation decision. If the binding information is validated as correct, then the PDF communicates the media authorisation details in the decision to the GGSN. The media authorisation details contain IP QoS parameters and packet classifiers related to the PDP context. The GGSN maps the authorised IP QoS parameters to authorised UTMS QoS parameters and, finally, the GGSN compares the UMTS QoS parameters with the authorised UMTS QoS parameters of the PDP context. If the UMTS QoS parameters from the PDP context request lie within the limits authorised by the PDF, then PDP context activation or modification will be accepted.

In addition to the bearer authorisation decision the PDF receives information about when an SBLP-governed PDP context is released, whether the UE has lost/recovered its radio bearer(s) and when an SBLP-governed PDP is using streaming or conversational traffic. Based on this information the PDF is able to inform the P-CSCF about the event that has occurred. This allows the P-CSCF to affect charging and it may even start to release an IMS session on behalf of the user. Moreover, the PDF is able to request the GGSN to deactivate a particular SBLP-governed PDP context.

The Security Gateway (SEG) has the function of protecting control plane traffic between security domains. A security domain refers to a network that is managed by a single administrative authority. Typically, this coincides with operator borders. The SEG is placed at the border of the security domain where it enforces the security policy of a security domain toward other SEGs in the destination security domain. In the IMS all traffic within the IMS is routed via SEGs, especially when the traffic is inter-domain, meaning that it originates in a different security domain from the one where it is received. When protecting inter-domain IMS traffic, both confidentiality as well as data integrity and authentication are mandated.

THIG functionality could be used to hide the configuration, capacity and topology of the network from outside an operator's network. If an operator wants to use a hiding functionality, then the operator must place a THIG function in the routing path when receiving requests or responses from other IMS networks. Similarly, the THIG must be placed in the routing path when sending requests or responses to other IMS networks. The THIG performs the encryption and decryption of all headers that reveal topology information about the operator's IMS network.

6.6.6 Charging Functions

The IMS architecture supports both online and offline charging capabilities. Online charging is a charging process in which IMS entities, such as an AS, interact with the online charging system. The online charging system in turn interacts in real time with the user's account and controls or monitors the charges related to service usage: for example, the AS queries the online charging system prior to allowing session establishment or it receives information about how long a user can participate in a conference. Offline charging is a charging process in which charging information is mainly collected after the session and the charging system does not affect in real time the service being used. In this model a user typically receives a bill on a monthly basis, which shows the chargeable items during a particular period. Due to the different nature of charging models different architecture solutions are required.

The central point in the offline charging architecture is the Charging Collection Function (CCF). The CCF receives accounting information from IMS entities (P-CSCF, S-CSCF, I-CSCF, BGCF, MGCF, AS, MRFC). It further processes the received data and then constructs and formats the actual Charging Data Record (CDR). The CDR is passed to the billing system, which takes care of providing the final CDR, taking into account information received from other sources as well (e.g., CGF). The billing system will create the actual bill. The bill could contain, for example, the number of sessions, destinations, duration and type of sessions (audio, text, video).

The S-CSCF, AS and MRFC are the IMS entities that are able to perform online charging. When the UE requests something from either the AS, the MRFC or the S-CSCF that requires charging authorisation, the entity contacts the Online Charging System (OCS) before delivering the service to the user: for example, the user could send a request to a news server asking for the latest betting odds or asking for a voice conference to be set up. The OCS supports two different authorisation models: immediate event charging and event charging with unit reservation. In the immediate event charging model the rating function of the OCS is used to find the appropriate tariff for an event. After resolving the tariff and the price, a suitable amount of money from the user's account is deducted and the OCS grants the request from the requesting entity (AS, MRFC or S-CSCF). When using this model the IMS entity should know that it could deliver the requested service to the user itself. For example, the AS could send a request and inform the OCS of the service (say, a game of chess) and the number of items (say, 2) to be delivered. Then the OCS uses the rating function to resolve the tariff (€0.3) and to calculate the price based on the number of delivered units (€0.6). Finally, €0.6 is deducted from the user's account and the OCS informs the AS that 2 units have been granted. In the event charging with unit reservation model the OCS uses the rating function to determine the price of the desired service according to service-specific information, if the cost was not given in the request. Then the OCS reserves a suitable amount of money from the user's account and returns the corresponding amount of resources to the requesting entity (AS, MRFC or S-CSCF). The amount of resources could be time or allowed data volume. When resources granted to the user have been consumed or the service has been successfully delivered or terminated, the IMS entity informs the OCS of the amount of resources consumed. Finally, the OCS deducts the amount used from the user's account [3GPP TS 32.200, TS 32.225, TS 32.260], but may require further interaction with the rating function. This model is suitable when the IMS entity (AS, MRFC or S-CSCF) is not able to determine beforehand whether the service could be delivered, or when the required amount of resources are not known prior to the use of a specific service (e.g., duration of the conference).